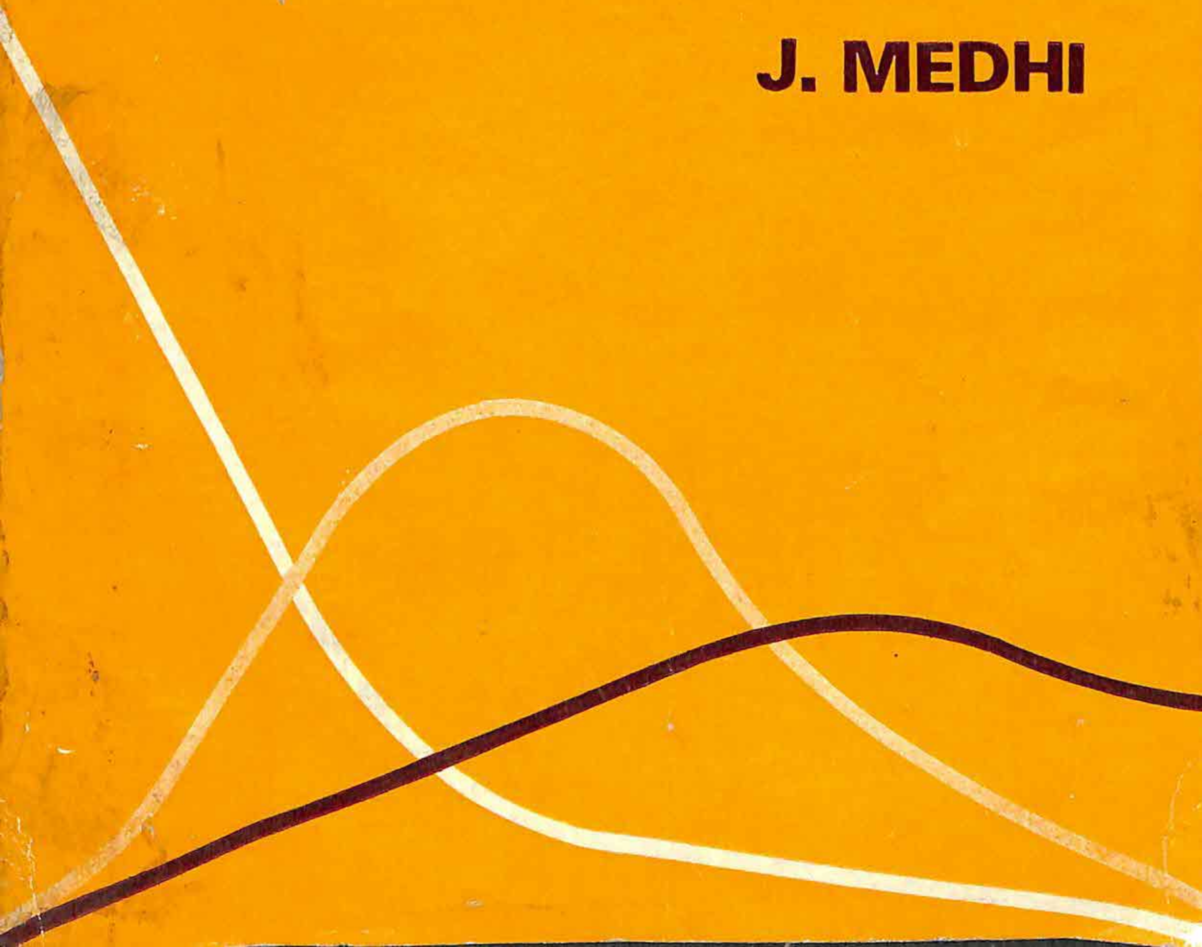
STOCHASTIC PROCESSES

J. MEDHI



3854
10.3.87

35/-

/c

STOCHASTIC PROCESSES

J. Medhi
Gauhati University
Gauhati

S.C.E.R.T., West Bengal
Date.........
Acc. No.............

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

WILEY EASTERN LIMITED
New Delhi Bangalore Bombay Calcutta Madras Hyderabad

S.C.E.R.T., West Bengal
Date 10-3-87
Acc. No. 3854

519.2
MED

Copyright ©1982, Wiley Eastern Limited

First Reprint, March 1983
Second Reprint, July 1984

WILEY EASTERN LIMITED

4835/24 Ansari Road, Daryaganj, New Delhi 110002
4654/21 Daryaganj, New Delhi 110002
6 Shri B.P. Wadia Road, Basavangudi, Bangalore 560004
Abid House, Dr Bhadkamkar Marg, Bombay 400 007
40/8 Ballygunge Circular Road, Calcutta 700 019
Post Box No. 1124, Tiruvanmiyur, Madras 600 041
Post Box No. 1050, Himayath Nagar P.O., Hyderabad 500 029

This book or any part thereof may not be reproduced in any form without the written permission of the publisher

This book is not to be sold outside the country to which it is consigned by Wiley Eastern Limited

10 9 8 7 6 5 4 3 2

ISBN 0 85226 564 6

Published by Mohinder Singh Sejwal for Wiley Eastern Limited 4835/24 Ansari Road, Daryaganj, New Delhi 110 002 and printed by Ravi Jain at Jupiter Offset Press, B-10/3, Jhilmil Industrial Area, Delhi 110032.

Printed in India.

Kalo hi eko nityaśca Vȳapakaḥ

—Bhāgavad-Purāṇa

To

My Mother

and to

The Memory of My Father

PREFACE

"Currently in the period of dynamic indeterminism in science, there is hardly a serious piece of research, which, if treated realistically, does not involve operations on stochastic processes" (Neyman, *J. Am. Stat. Ass.* (1960) p. 625). It is thus natural to find a growing awareness and interest in this subject in scientific and technological studies everywhere.

It is felt that, unlike statistical theory, a large variety of introductory and intermediate level texts dedicated to stochastic processes are not available, and that there is need for more books at this level. This book which aims at the level between that of elementary probability texts and of the excellent advanced works on stochastic processes, has been written keeping in view students of diverse interests and background. The book deals mainly with methods and applications.

The prerequisites for using this book are a course on elementary probability theory and statistics and a course on advanced calculus. Some of the basic mathematical concepts and techniques used (in the sequel) have been treated in some details in Chapter 1. The theoretical results developed have been followed by a large number of illustrative examples. These have been supplemented by a large number of exercises, answers to most of which are also given. Lists of references are provided for the benefit of those who need more details or wish to pursue the subject further.

It is expected that the book would be suitable as a text book for advanced undergraduate, post graduate and research level courses in Statistics, Applied Mathematics, Operations Research, Computer Science, Management Science etc. as also for graduate and beginning research level courses of study in Physics, Life Sciences and different branches of Engineering. The course could be covered in one or two semesters depending on the interest of the users and on the extent or coverage desired. Researchers who need knowledge of applied probability and applied stochastic processes would also profit from using the book. The illustrative examples spread over areas, such as, biological sciences to management sciences would lend a special appeal to not only those who wish to pursue the subject but also to those who wish to go for applications. Some materials, specially on applications from some of the most recent papers are also included. There is a somewhat exhaustive chapter on stochastic processes in Queues, and this portion could be used also for a course on queueing theory. The material presented in the book would, it is hoped, be suitable, in general, for students having some mathematical maturity who wish to gain an understanding of the subject and to see how the theory works.

The book has grown out of the author's long 30 years' experience of teaching and research. The suggestion of writing a book on the lines of his lecture notes, was first made to him by groups of students who took such a course given by him at the University of Montreal, Canada.

I look forward to receive comments and suggestions on the work from students, teachers and researchers.

I remain intellectually indebted to all those workers whose works have stimulated my interest in the subject. In the present venture, encouragements and suggestions have been received from a number of learned scholars both in this country and abroad. I am thankful to them all.

I am indeed most grateful to Professor N.U. Prabhu, School of Operations Research, Cornell University, U.S.A. and to Professor R. Fortet, Université de Paris, France, both of whom kindly read the manuscript thoroughly and made very valuable comments and offered constructive suggestions. I recall also the appreciative interest shown and the constant encouragement given by Professor Prabhu.

I am indebted to Professor M.B. Priestley, University of Manchester Institute of Science & Technology, U.K. and to Professor R.M. Loynes, University of Sheffield, U.K. for having glanced through the manuscript within a short time and for making useful comments, specially with regard to coverage.

I received considerable assistance with proofreading from some of my former research students and present colleagues, namely, Dr S.B. Nandi, Dr. A. Borthakur, Dr. H.K. Baruah and Mr. D.C. Jain as well as from my elder son, Deepankar; and help in many other ways from my younger son, Shubhankar and two daughters, Shakuntala and Alakanandaa. Acknowledgement is also made of the excellent typing by S. Borah and A. Choudhury.

JYOTIPRASAD MEDHI

Gauhati, India
August 1981

CONTENTS

Preface vii

1. INTRODUCTION 1

1.1 Generating Functions *1*
1.2 Laplace Transforms *12*
1.3 Laplace (Stieltjes) Transform of a Probability Distribution or of a Random Variable *15*
1.4 Difference Equations 24
1.5 Difference Equations in Probability Theory *27*
1.6 Differential-Difference Equations *30*
1.7 Matrix Analysis *33*

2. STOCHASTIC PROCESSES 45

2.1 Notion of Stochastic Processes *45*
2.2 Specification of Stochastic Processes *46*
2.3 Stationary Processes *48*

3. MARKOV CHAINS 54

3.1 Definition and Examples *54*
3.2 Higher Transition Probabilities *60*
3.3 Classification of States and Chains *63*
3.4 Determination of Higher Transition Probabilities *69*
3.5 Stability of a Markov System: Limiting Behaviour *73*
3.6 Statistical Inference for Markov Chains *79*
3.7 Markov Chains with Continuous State Space *84*
3.8 Non-stationary or Non-homogeneous Chains *87*

4. MARKOV PROCESSES WITH DISCRETE STATE SPACE: POISSON PROCESS AND ITS EXTENSIONS 96

4.1 Poisson Process *96*
4.2 Poisson Process and Related Distributions *107*
4.3 Generalisations of Poisson Process *113*
4.4 Birth and Death Process *120*
4.5 Markov Processes with Discrete State Space (Continuous Time Markov Chains) *126*
4.6 Erlang Process *138*

5. MARKOV PROCESSES WITH CONTINUOUS STATE SPACE 146

5.1 Introduction: Brownian Motion *146*

5.2 Wiener Process *147*
5.3 Differential Equations for a Wiener Process *149*
5.4 Kolmogorov Equations *150*
5.5 First Passage Time Distribution for Wiener Process *152*
5.6 Ornstein-Uhlenbeck Process *156*

6. RENEWAL PROCESSES AND THEORY 164

6.1 Renewal Processes *164*
6.2 Renewal Equation *167*
6.3 Renewal Theorems *169*
6.4 Delayed and Equilibrium Renewal Processes *177*
6.5 Residual and Excess Lifetimes *186*
6.6 Renewal Reward Process *190*
6.7 Alternating (or Two-Stage) Renewal Process *192*
6.8 Regenerative Stochastic Process: Existence of Limits *195*

7. MARKOV RENEWAL AND SEMI-MARKOV PROCESSES 203

7.1 Introduction *203*
7.2 Definitions and Preliminary Results *204*
7.3 Markov Renewal Equation *210*
7.4 Limiting Behaviour *212*
7.5 First Passage Time *216*

8. STATIONARY PROCESSES AND TIME SERIES 223

8.1 Introduction *223*
8.2 Models of Time Series *224*
8.3 Time and Frequency Domain: Power Spectrum *230*
8.4 Statistical Analysis of Time Series: Some Observations *235*

9. BRANCHING PROCESSES 241

9.1 Introduction *241*
9.2 Properties of Generating Functions of Branching Processes *242*
9.3 Probability of Extinction *244*
9.4 Distribution of Total Number of Progeny *252*
9.5 Conditional Limit Laws *253*
9.6 Continuous Time Markov Branching Processes *262*
9.7 Age-dependent Branching Process: Bellman-Harris Process *267*

10. STOCHASTIC PROCESSES IN QUEUEING AND RELIABILITY 277

10.1 Queueing Systems: General Concepts *277*
10.2 The Queueing Model *M*/*M*/1: Steady State Behaviour *280*
10.3 Transient Behaviour of *M*/*M*/1 Model *286*

10.4 Birth and Death Processes in Queueing Theory: Multi-channel Models *294*
10.5 Non-Birth and Death Queueing Processes: Bulk Queues *303*
10.6 The Queueing Models *M/G/*1 and *GI/M/*1 *311*
10.7 Reliability *331*

Answers To Exercises **356**

Appendix **363**

1. Abbreviations
   Glossary of Commonly Used Notations
2. Table of Laplace Transforms
3. Important Probability Distributions and Their Parameters, Moments, Generating Functions

Subject Index **377**

Author Index **385**

CHAPTER 1

INTRODUCTION

1.1 Generating Functions

A. Introduction

In dealing with integral-valued random variables, it is often of great convenience to apply the powerful method of generating functions (Feller, Vol. I, Chapter XI). Many stochastic processes that we come across involve integral-valued random variables, and quite often we could use generating function in their studies. The principal advantage of its use is that a single function (generating function) may be used to represent a whole set of individual items.

Definition. Let $a_0, a_1, a_2, \ldots$ be a sequence of real numbers. Using a variable s, we may define a function

$$A(s) = a_0 + a_1 s + a_2 s^2 + \ldots = \sum_{k=0}^{\infty} a_k s^k. \tag{1.1}$$

If this power series converges in some interval $-s_0 < s < s_0$, then $A(s)$ is called the generating function of the sequence $a_0, a_1, a_2, \ldots$. The variable s itself has no particular significance. Here we assume s to be real but a generating function with a complex variable is also sometimes used.

Differentiating (1.1) k times, putting $s = 0$ and dividing by $k!$, we get a_k, i.e.

$$a_k = \frac{1}{k\,!}\left[\frac{d^k A}{ds^k}\right]_{s=0}. \tag{1.2}$$

B. Probability Generating Function : Mean and Variance

Suppose that X is a random variable which assumes non-negative integral values 0, 1, 2, ..., and that

$$\Pr\{X = k\} = p_k,\ k = 0, 1, 2, \ldots, \sum_k p_k = 1. \tag{1.3}$$

If we take a_k to be the probability p_k, $k = 0, 1, 2, \ldots$ then the corresponding generating function $P(s)$ of the sequence of probabilities $\{p_k\}$ is known as the *probability generating function* (p.g.f.) of the random variable X. It is sometimes also called the *s-transform* (or geometric transform) of X.

We have

$$P(s) = \sum_{k=0}^{\infty} p_k s^k \tag{1.4}$$

$$= \sum_{k=0}^{\infty} \Pr\{X = k\}\, s^k = E(s^X), \tag{1.4a}$$

where $E(s^X)$ is the expectation of the function s^X (a random variable) of the random variable X.

The series $P(s)$ converges for at least $-1 < s \leqslant 1$; clearly $P(1) = 1$. The first two derivatives of $P(s)$ are given by

$$P'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1}, \quad -1 < s < 1 \tag{1.5a}$$

and

$$P''(s) = \sum_{k=1}^{\infty} k(k-1)\, p_k s^{k-2}. \tag{1.5b}$$

The expectation $E(X)$ is given by

$$E(X) = \sum_{k=1}^{\infty} k p_k = P'(1). \tag{1.6}$$

We have

$$E\{X(X-1)\} = \sum_{k} k(k-1)\, p_k = P''(1) \tag{1.7}$$

and

$$E(X^2) = E\{X(X-1)\} + E(X) = P''(1) + P'(1).$$

Hence

$$\begin{aligned} \operatorname{var}(X) &= E(X^2) - [E(X)]^2 \\ &= P''(1) + P'(1) - [P'(1)]^2. \end{aligned} \tag{1.8}$$

If, as $s \to 1$, $\Sigma\, k p_k$ diverges then we say that $E(X) = P'(1) = \infty$ and if $\Sigma\, k(k-1)\, p_k$ diverges then $P''(1) = \infty$ and $\operatorname{var}(X) = \infty$.

Equation (1.6) gives the mean and (1.8) the variance of X in terms of the p.g.f. of X. In fact, moments, cumulants etc. can be expressed in terms of generating functions. More generally, the kth factorial moment of X is given by

$$E\{X(X-1)\dots(X-k+1)\} = \frac{d^k}{ds^k} P(s)\Big|_{s=1} \qquad \text{for } k = 1, 2, \dots$$

Note that $P(e^t)$ is the moment generating function and $P(1+t)$ is the factorial moment generating function. Further, the p.g.f. $P(s) = E(s^X)$ is a special case of the characteristic function $E(e^{itX})$ or the Fourier transform of X.

Note also that $$\int_0^1 P(s)\, ds = E\left(\frac{1}{X+1}\right).$$

Example 1(a). The sequence $\{a_k = (1/k!)\}$ has the generating function

$$A(s) = \sum_{k=0}^{\infty} a_k s^k = \sum_{k=0}^{\infty} \frac{s^k}{k\,!} = e^s.$$

Here $\{a_k\}$ is a sequence of real numbers. This does not represent probabilities of any random variable.

Example 1(b). *Poisson distribution*: Let X be a Poisson variate with probability function

$$p_k = \Pr\{X = k\} = \frac{\exp(-\lambda)\,\lambda^k}{k\,!}, \quad k = 0, 1, 2, \ldots$$

The p.g.f. of Poisson distribution (or of the random variable X) is given by

$$\begin{aligned} P(s) = \sum_{k=0}^{\infty} p_k s^k &= \sum_{k=0}^{\infty} \frac{\exp(-\lambda)\,\lambda^k}{k\,!} s^k \\ &= \exp(-\lambda)\exp(\lambda s) \\ &= \exp\{\lambda(s-1)\}. \end{aligned}$$

Since $P'(s) = \lambda \exp\{\lambda(s-1)\}$ and $P''(s) = \lambda^2 \exp\{\lambda(s-1)\}$, we have, from (1.6) and (1.8)

$$E(X) = \lambda \quad \text{and} \quad \operatorname{var}(X) = \lambda.$$

Example 1(c). *Geometric distribution*: Let X be a random variable with geometric distribution $\{p_k\}$, where

$$p_k = \Pr\{X = k\} = q^k p, \qquad k = 0, 1, 2, \ldots$$

$$p + q = 1.$$

The p.g.f. of X is

$$P(s) = \sum_{k=0}^{\infty} p_k s^k = p \sum_{k=0}^{\infty} (qs)^k = \frac{p}{1-qs}.$$

We have

$$P'(s) = \frac{pq}{(1-qs)^2}, \quad P''(s) = \frac{2pq^2}{(1-qs)^2}.$$

From (1.6) and (1.8) we have

$$E(X) = P'(1) = \frac{pq}{(1-q)^2} = \frac{q}{p}$$

and

$$\operatorname{var}(X) = P''(1) + P'(1) - [P'(1)]^2 = \frac{2q^2}{p^2} + \frac{q}{p} - \left(\frac{q}{p}\right)^2 = \frac{q}{p^2}.$$

An example of a random variable X which follows geometric distribution is given below. Consider a sequence of Bernoulli trials (i.e. independent trials with only two possible outcomes "success" and "failure" at each trial, the probability of success being p and that of failure being $q = 1 - p$). Then the number X of failures preceding the first success has geometric distribution with $\Pr\{X=k\}=q^k p\,(k=0,1,2,\ldots)$. X is also called the *waiting time* for the first success. Note that the continuous analogue of the geometric distribution is the exponential distribution.

Example 1(d). *Binomial distribution:* Let X be a random variable with binomial distribution (number of successes in n Bernoulli trials) with

$$p_k = \Pr\{X=k\} = \binom{n}{k} p^k q^{n-k},\quad k=0,1,2,\ldots,n.$$

$$p+q=1.$$

The p.g.f. of X is

$$P(s)=\sum_{k=0}^{\infty} p_k s^k = \sum_{k=0}^{n}\binom{n}{k} p^k q^{n-k} s^k = (q+sp)^n.$$

Thus

$$E(X) = P'(1) = np$$

and

$$\begin{aligned}\operatorname{var}(X) &= P''(1)+P'(1)-[P'(1)]^2\\ &= n(n-1)p^2+np-n^2p^2\\ &= npq.\end{aligned}$$

Example 1(e). Let X be a random variable with p.g.f. $P(s)$. To find the p.g.f. of the random variable $Y=mX+n$, where m, n are integers and $m\neq 0$.

Let

$$P_X(s) \text{ and } P_Y(s)$$

be the generating functions of X and Y respectively. We have

$$\begin{aligned}P_Y(s) &= E\{s^Y\} = E\{s^{mX+n}\}\\ &= E\{s^{mX}.s^n\} = s^n.E\{(s^m)^X\}\\ &= s^n P_X(s^m).\end{aligned}$$

From the above examples, we see how a single generating function $P(s)$ may be used to represent a whole set of probabilities

$$p_k = \Pr\{X=k\},\ k = 0,1,2,\ldots.$$

In these examples we were concerned with the problem of finding $P(s)$ for a given set of p_k's.

In many cases the reverse problem arises: to determine p_k from a given p.g.f. $P(s)$. Many situations arise, where it is easier to find the p.g.f. $P(s)$ of a variable rather than the probability distribution $\{p_k\}$ of the variable. One proceeds to find first the p.g.f. $P(s)$ and then to find the probabilities p_k from the function $P(s)$. Even without finding the p_k's one can find the moments of the distribution from $P(s)$.

Determination of p_k from $P(s)$: p_k can be (uniquely) determined from $P(s)$ as follows:

p_k can be found from $P(s)$ by applying (1.2), i.e.

$$p_k = \frac{1}{k!}\left[\frac{d^k P(s)}{ds^k}\right]_{s=0}.$$

p_k is also given by the coefficient of s^k in the expansion of $P(s)$ as a power series in s.

When $P(s)$ is of the form $P(s) = U(s)/V(s)$, it may be convenient to expand $P(s)$ in a power series in s first by decomposing $P(s)$ into partial fractions. Suppose that $s_1, \ldots, s_r$ are the distinct roots of $V(s)$, i.e. $V(s) = (s - s_1)\ldots(s - s_r)$ (apart from a constant factor c which, for simplicity, we take to be equal to 1), then $P(s)$ can be decomposed into partial fractions as

$$P(s) = \frac{a_1}{s_1 - s} + \ldots + \frac{a_r}{s_r - s},$$

where a's can be determined. It may be verified that

$$a_i = -U(s_i)/V'(s_i).$$

Now

$$\begin{aligned}\frac{a_i}{s_i - s} &= \frac{a_i}{s_i(1 - s/s_i)} = \frac{a_i}{s_i}\left(1 - \frac{s}{s_i}\right)^{-1} \\ &= \frac{a_i}{s_i}\left\{1 + \frac{s}{s_i} + \left(\frac{s}{s_i}\right)^2 + \ldots\right\} \quad \text{for } |s| < |s_i| \\ &= a_i \sum_{k=0}^{\infty} \frac{s^k}{s_i^{k+1}}.\end{aligned}$$

Hence $p_k \equiv$ coefficient of s^k in $P(s)$

$$= \frac{a_1}{s_1^{k+1}} + \ldots + \frac{a_i}{s_i^{k+1}} + \ldots + \frac{a_r}{s_r^{k+1}}.$$

C. Sum of a fixed number of Random Variables

Let X and Y be two independent non-negative integral-valued random variables with probability distributions given by

$$\Pr\{X = k\} = a_k, \quad \Pr\{Y = j\} = b_j.$$

The sum $Z = X + Y$ is a random variable. The event $\{Z = r\}$ can happen in the following mutually exclusive ways with corresponding probabilities:

$(X = 0$ and $Y = r)$ with probability $a_0 b_r$

$(X = 1$ and $Y = r - 1)$ with probability $a_1 b_{r-1}$

... ...

$(X = t$ and $Y = r - t)$ with probability $a_t b_{r-t}$

... ...

$(X = r$ and $Y = 0)$ with probability $a_r b_0$.

Hence the distribution of Z is given by

$$c_r = \Pr\{Z = r\} = a_0 b_r + a_1 b_{r-1} + \ldots + a_r b_0 = \sum_{t=0}^{r} a_t b_{r-t}. \qquad (1.9)$$

This new sequence $\{c_r\}$ which results from a combination of the two sequences $\{a_k\}$, $\{b_j\}$ in this particular way indicated by (1.9) is called the *convolution* of $\{a_k\}$ and $\{b_j\}$ and is denoted by

$$\{c_r\} = \{a_r\} * \{b_r\}.$$

Let $A(s)$, $B(s)$, $C(s)$ be the p.g.f. of X, Y and Z respectively. Then from (1.9) it follows that

$$C(s) = A(s) \cdot B(s). \qquad (1.9a)$$

This also follows from the fact that

$$\begin{aligned} C(s) &= E(s^{X+Y}) = E(s^X)\, E(s^Y), \\ &= A(s) \cdot B(s), \end{aligned}$$

because of independence of X, Y.

We have thus the following theorem:

Theorem 1.1. *The p.g.f. of the sum of two independent random variables X and Y is the product of the p.g.f. of X and that of Y.*

The result will also hold good in the case of the sum S_n of n non-negative independent integral-valued random variables $X_1, \ldots, X_n$, i.e. the p.g.f. of S_n is the product of the p.g.f.'s of $X_1, X_2, \ldots, X_n$.

In particular, we have the following:

Theorem 1.2. The sum $S_n = X_1 + \ldots + X_n$ of a *fixed* number n of identically and independently distributed (i.i.d.) random variables X_i has the p.g.f. $\{P(s)\}^n$, each X_i having p.g.f. $P(s)$.

The results stated above are of considerable importance. These can be employed to find the distribution of the sum of a number of independent random variables.

Example 1(f). *Sum of independent Poisson variates*: Let X_1, X_2 be two independent Poisson variates with parameters (means) λ_1, λ_2 respectively. To find the distribution of the sum $Z = X_1 + X_2$.

The p.g.f. of X is $\exp\{\lambda_i(s-1)\}$, $i = 1, 2$.

Hence the p.g.f. of Z is

$$\begin{aligned} &[\exp\{\lambda_1(s-1)\}]\,[\exp\{\lambda_2(s-1)\}] \\ &= \exp\{(\lambda_1+\lambda_2)(s-1)\}. \end{aligned}$$

But this is the p.g.f. of a Poisson variate with parameter $\lambda_1 + \lambda_2$. Hence the sum of two independent Poisson variates is a Poisson variate. More generally, if X_i, $i = 1, 2, \ldots, n$ are n independent Poisson variates with parameters λ_i, then the sum $S = X_1 + X_2 + \ldots + X_n$ is a Poisson variate with parameter $\sum_{i=1}^{n} \lambda_i$.

Example 1(g). Let X_1, X_2 be two independent random variables each with the same geometric distribution, $\Pr\{X_i = k\} = q^k p$, $i = 1, 2$.

The p.g.f. of X_i is $p/(1 - qs)$ for $i = 1, 2$. Thus, the sum $Z = X_1 + X_2$ has the p.g.f.

$$P(s) = \left(\frac{p}{1 - qs}\right)^2 = p^2 \sum_{r=0}^{\infty} \binom{r+1}{r} q^r s^r.$$

Hence,

$$\begin{aligned}\Pr\{Z = r\} &\equiv \text{coefficient of } s^r \text{ in } P(s)\\ &= \binom{r+1}{r} q^r p^2 = (r+1)\, q^r p^2,\ r = 0, 1, 2, \ldots.\end{aligned}$$

This gives the distribution of the sum Z. If X_1 denotes the number of failures preceding the first success and X_2 the number of failures following the first success and preceding the second success, then $Z = X_1 + X_2$ denotes the number of failures preceding the second success.

The sum $S_r = X_1 + \ldots + X_r$ of r i.i.d. geometric variates denotes the number of failures preceding the rth success. It has the p.g.f. $P(s) = \left(\frac{p}{1 - sq}\right)^r$. It follows that

$$\begin{aligned}\Pr\{S_r = k\} &\equiv \text{coefficient of } s^k \text{ in } P(s)\\ &= \binom{r+k-1}{k} p^r q^k, \quad k = 0, 1, 2, \ldots \qquad (1.10)\end{aligned}$$

and

$$\sum_k \binom{r+k-1}{k} p^r q^k = 1 \qquad (1.10a)$$

Here r is a positive integer. But the quantity given by the r.h.s. of (1.10) is non-negative and (1.10a) holds for any positive r. The distribution (1.10) for $r > 0$ is known as *negative binomial distribution* with *index* r, and mean rq/p. In particular, when r is a positive integer, it is known as *Pascal distribution*.

Example 1(h). If X, Y are independent r.v.'s. with generating functions $A(s)$, $B(s)$ respectively, then the generating function $D(s)$ of $W = X - Y$ is given by

$$D(s) = E\{s^{X-Y}\} = E\{s^X (s^{-1})^Y\} = E\{s^X\}\, E\{(1/s)^Y\},$$

because of independence of X, Y; thus

$$D(s) = A(s)\, B(1/s).$$

D. Sum of a Random Number of Discrete Random Variables

In Section C we considered the sum $S_n = X_1 + \ldots + X_n$ of a fixed number of mutually independent random variables. Sometimes we come across situations when we have to consider the sum of *a random number N* of random variables. For example, if X_i denotes the number of persons involved in the ith accident (in a day in a certain city) and if the number

N of accidents happening on a day is a random variable, then the sum $S_N = X_1 + \ldots + X_N$ denotes the total number of persons involved in accidents on a day.

We have the following theorem for the sum S_N.

Theorem 1.3. Let X_i, $i = 1, 2, \ldots$, be identically and independently distributed (i.i.d.) random variables with $\Pr\{X_i = k\} = p_k$ and p.g.f.

$$P(s) = \sum_k p_k s^k \quad \text{for } i = 1, 2, \ldots \tag{1.11}$$

Let

$$S_N = X_1 + \ldots + X_N, \tag{1.12}$$

where N is a random variable independent of the X_i's. Let the distribution of N be given by $\Pr\{N = n\} = g_n$ and the p.g.f. of N be

$$G(s) = \sum g_n s^n. \tag{1.13}$$

Then the p.g.f. $H(s)$ of S_N is given by the compound function $G(P(s))$, i.e.

$$H(s) = \sum_j \Pr\{S_N = j\}\, s^j = G(P(s)). \tag{1.14}$$

Proof: Since N is a random variable which can assume values 0,1,2,3,..., the event $S_N = j$ can happen in the following mutually exclusive ways: $N = n$ and $S_n = X_1 + \ldots + X_n = j$ for $n = 1, 2, 3, \ldots$. To meet the situation $N = 0$, let us define $X_0 = S_0 = 0$ so that

$$X_0 + X_1 + \ldots + X_n = S_n, \ n \geqslant 0.$$

We have $\quad h_j = \Pr\{S_N = j\} = \sum_{n=0}^{\infty} \Pr\{N = n \text{ and } S_n = j\}.$

Since N is independent of X_i's and therefore of S_n,

$$h_j = \sum_{n=0}^{\infty} \Pr\{N = n\} \Pr\{S_n = j\} \tag{1.15}$$

$$= \sum_{n=0}^{\infty} g_n \Pr\{S_n = j\}. \tag{1.16}$$

(The inclusion of the value 0 for n can be justified since $\Pr\{S_0 = 0\} = 1$ and $\Pr\{S_0 > 0\} = 0$.)

The p.g.f. of the sum $S_N = X_1 + \ldots + X_N$ is thus given by

$$H(s) = \sum_j h_j s^j = \sum_j \Pr\{S_N = j\}\, s^j \tag{1.17}$$

$$= \sum_{j=0}^{\infty} \Big[\sum_{n=0}^{\infty} g_n \Pr\{S_n = j\}\Big] s^j \tag{1.18}$$

$$= \sum_n \Big[\sum_j \Pr\{S_n = j\}\, s^j\Big] g_n$$

$$= \sum_n g_n [P(s)]^n = G(P(s)),$$

since the expression $[P(s)]^n$ is the p.g.f. of the sum S_n of a fixed number n of i.i.d. random variables (Theorem 1.2). ▲

Corollary: The mean

$$E\{S_N\} = E\{X_i\}\, E\{N\}. \tag{1.19}$$

We have
$$\begin{aligned} E\{S_N\} &= H'(s)\,|_{s=1} = [P'(s)G'(P(s))]_{s=1} \\ &= P'(1)\, G'(P(1)) = P'(1)\, G'(1) \\ &= E\{X_i\}\, E\{N\}. \end{aligned}$$

Note. If $P(s)$ and $G(s)$ are two p.g f.'s, then $G(P(s))$ is also a p.g.f.

Example 1(i). *Compound Poisson distribution:* If N has a Poisson distribution with mean λ, then $G(s) = \exp\{\lambda(s-1)\}$ and hence the sum $X_1 + \ldots + X_N$ has p.g.f.

$$H(s) = G(P(s)) = \exp\{\lambda\,[P(s) - 1]\}. \tag{1.20}$$

The distribution having a generating function of the form $\exp\{\lambda[P(s) - 1]\}$, where $P(s)$ is itself a generating function is called *compound Poisson* distribution.

The mean is given by

$$E\{S_N\} = E\{X_i\}\, E\{N\} = \lambda\, E(X_i).$$

E. Generating Function of Bivariate Distribution

Bivariate Distribution: Suppose that X, Y is a pair of integral-valued random variables with joint probability distribution given by

$$\Pr\{X = j,\ Y = k\} = p_{jk},\ j, k = 0, 1, 2, \ldots, \sum_{j,k} p_{jk} = 1. \tag{1.21}$$

The marginal distributions are given by

$$\Pr\{X = j\} = \sum_{k=0}^{\infty} p_{jk} = f_j, \qquad j = 0, 1, 2, \ldots \tag{1.22}$$

$$\Pr\{Y = k\} = \sum_{j=0}^{\infty} p_{jk} = g_k, \qquad k = 0, 1, 2, \ldots. \tag{1.23}$$

If (j, k) is any point at which $f_j \neq 0$ (i.e. $f_j > 0$), then the conditional distribution of Y given $X = j$ is given by

$$\Pr\{Y = k \mid X = j\} = \frac{\Pr\{Y = k,\ X = j\}}{\Pr\{X = j\}} = \frac{p_{jk}}{f_j}, \quad k = 0, 1, 2, \ldots. \tag{1.24}$$

Conditional Expectation

The conditional expectation $E\{Y \mid X = j\}$ is given by

$$\begin{aligned} E\{Y \mid X = j\} &= \sum_k k \Pr\{Y = k \mid X = j\} \\ &= \sum_k k \frac{p_{jk}}{f_j} = \frac{\sum_k k p_{jk}}{f_j}, \quad j = 0, 1, 2, \ldots. \end{aligned} \tag{1.25}$$

For X given but unspecified, we note that $E\{Y \mid X\}$ is a random variable which assumes the value $\dfrac{\sum_k kp_{jk}}{f_j}$ with $\Pr\{X=j\}=f_j>0$.

Hence the expectation of the random variable $E\{Y \mid X\}$ is

$$\begin{aligned} E[E\{Y \mid X\}] &= \sum_j E\{Y \mid X=j\} \Pr\{X=j\} \\ &= \sum_j \left\{\frac{\sum_k kp_{jk}}{f_j}\right\} f_j \\ &= \sum_j \sum_k kp_{jk} \\ &= \sum_k k\{\sum_j p_{jk}\} = \sum_k k \Pr\{Y=k\} = E(Y). \end{aligned} \tag{1.26}$$

In the same way we can prove that

$$E[E\{Y^2 \mid X\}] = E[Y^2]; \tag{1.27}$$

more generally, for any function $\phi(Y)$ whose expectation exists,

$$E[E\{\phi(Y) \mid X\}] = E[\phi(Y)]. \tag{1.28}$$

Note: The results (1.26–1.28) which are given here for discrete random variables X, Y will also hold, *mutatis mutandis*, for continuous random variables.

Bivariate Probability Generating Function

Definition. The probability generating function (bivariate p.g.f.) of a pair of random variables X, Y with joint distribution given by (1.21) is defined by

$$P(s_1, s_2) = \sum_{j,k} p_{jk} s_1{}^j s_2{}^k, \tag{1.29}$$

s_1, s_2 being dummy positive real variables chosen so as to make the double series convergent. We have the following result.

Theorem 1.4. (a) The p.g.f. $A(s)$ of the marginal distribution $\Pr\{X=j\}$ is given by $A(s)=P(s, 1)$.

(b) The p.g.f. $B(s)$ of $\Pr\{Y=k\}$ is given by

$$B(s) = P(1, s).$$

(c) The p.g.f. of $(X+Y)$ is given by $P(s, s)$.

Proof: Since the convergent double series (1.29) consists of positive real terms, the change of order of summation is justified.

(a) We have from (1.29)

$$P(s, 1) = \sum_{j,k} p_{jk} s^j = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} p_{jk} s^j$$

$$= \sum_{j=0}^{\infty} \{ \sum_{k=0}^{\infty} p_{jk} \} s^j = \sum_{j=0}^{\infty} \Pr\{X=j\} s^j, \text{from } (1.22)$$

$$= A(s), \text{ which is the p.g.f. of } X.$$

(b) can be proved in the same way.

(c) We have

$$P(s, s) = \sum_{j,\, k=0}^{\infty} p_{jk} s^{j+k} = \sum_{m=0}^{\infty} \sum_{r=0}^{m} p_{r\ m-r} s^m$$

as can be easily verified.

Now $\Pr\{X+Y=m\} = \sum_{r=0}^{m} \Pr\{X=r,\ Y=m-r\} = \sum_{r=0}^{m} p_{r\ m-r}.$

Hence

$$P(s, s) = \sum_{m=0}^{\infty} [\Pr\{X+Y=m\}]\, s^m,$$

which shows that $P(s, s)$ is the p.g.f. of $X+Y$. ▲

Remarks: **(1)** If X, Y are independent,

$$p_{jk} = \Pr\{X=j,\ Y=k\} = \Pr\{X=j\} \Pr\{Y=k\}$$

and then $\qquad P(s_1, s_2) = A(s_1)\, B(s_2) = P(s_1, 1)\, P(1, s_2)$

and conversely, from this relation follows the independence of X and Y.

(2) The probabilities p_{jk} can be (uniquely) determined from $P(s_1, s_2)$ as follows:

$$p_{jk} = \frac{1}{j!\,k!}\left[\frac{\partial^{j+k} P(s_1, s_2)}{\partial s_1{}^j\, \partial s_2{}^k}\right]_{s_1=s_2=0}$$

Example 1(j). Consider a series of Bernoulli trials with probability of success p. Suppose that X denotes the number of failures preceding the first success and Y the number of failures following the first success and preceding the second success.

The joint distribution of X, Y is given by

$$p_{jk} = \Pr\{X=j,\ Y=k\} = q^{j+k}p^2, \qquad j, k = 0, 1, 2, \ldots;$$

the bivariate generating function is

$$P(s_1, s_2) = \sum_{j,k=0}^{\infty} p_{jk}\, s_1^j s_2^k = \sum_{j,k}^{\infty} q^{j+k}\, p^2\, s_1^j s_2^k$$

$$= p^2 \{ \sum_{j=0}^{\infty} (qs_1)^j \} \{ \sum_{k=0}^{\infty} (qs_2)^k \}$$

$$= \frac{p^2}{(1-s_1 q)(1-s_2 q)}.$$

The p.g.f. of $\{X\}$ is given by

$$A(s) = P(s, 1) = \frac{p^2}{(1-sq)(1-q)} = \frac{p}{1-qs}.$$

The p.g.f. of $X + Y$ is given by

$$P(s,s) = \left(\frac{p}{1 - sq}\right)^2.$$

1.2 Laplace Transforms

A. Introduction

Laplace transform is a generalisation of generating function. Laplace transforms serve as a very powerful tool in many situations. They provide an effective means for the solution of many problems arising in our study. For example, for the solution of difference equations the transforms provide a very effective method. In the study of some probability distributions, the method could be used with great advantage, for it happens quite often that it is easier to find the Laplace transform of a probability distribution rather than the distribution itself. Here we shall mention some of the important properties of the Laplace transforms and illustrate the method by means of some examples (see Widder).

Definition. Let $f(t)$ be a function of a positive real variable t. Then the Laplace transform (L.T.) of $f(t)$ is defined by

$$F(s) = \int_0^\infty \exp(-st) f(t)\, dt \tag{2.1}$$

for the range of value of s for which the integral exists.

We shall write $F(s) = L\{f(t)\}$ to denote the Laplace transform of $f(t)$.

Example 2(a).

(i) If $f(t) = 1$, $F(s) = \int_0^\infty \exp(-st)\, dt = 1/s \quad (s > 0)$

If $f(t) = t$, $F(s) = \int_0^\infty t \exp(-st)\, dt = 1/s^2 \quad (s > 0)$

If $f(t) = t^n (n > -1)$, $F(s) = \int_0^\infty t^n \exp(-st)\, dt = \Gamma(n + 1)/s^{n+1}$

$(s > 0)$

(ii) Let $f(t) = e^{at}$, then $F(s) = \int_0^\infty \exp(-st) \exp(at)\, dt$

$= 1/(s - a) \quad (s > a)$

(iii) Let $f(t) = \sin t$, then
$$F(s) = \int_0^\infty \exp(-st) \sin t \, dt$$
$$= 1/(s^2 + 1) \qquad (s > 0)$$

(iv) Let $f(t) = e^{-t} t^n$, then
$$F(s) = \int_0^\infty \exp(-st) \exp(-t) \, t^n \, dt$$
$$= \int_0^\infty \exp\{-t(s+1)\} \, t^n \, dt$$
$$= \frac{\Gamma(n+1)}{(s+1)^{n+1}} \qquad (n > -1).$$

B. Some Important Properties of Laplace Transforms

We state the results without proof; the reader may refer to any standard work for proof.

1. *Linearity property*: If c_i's are constants, and
$$L\{f_i(t)\} = F_i(s), \; i = 1, 2, \ldots n,$$
then
$$L\{c_1 f_1(t) + \ldots + c_n f_n(t)\} = c_1 F_1(s) + \ldots + c_n F_n(s), \qquad (2.2)$$
e.g.
$$L\{2t - 5\} = 2/s^2 - 5/s.$$

2. *First translation property*:
$$L\{(\exp(-at)) f(t)\} = F(s + a), \qquad a > 0, \qquad (2.3)$$
e.g.
$$L\{(\exp(-t)) \, t\} = \frac{1}{(s+1)^2}.$$

(2.3) holds also for such negative values of a for which the corresponding integral exists.

3. *Second translation property*:

If
$$G(t) = \begin{cases} f(t-a), & t > a \\ 0 & t < a, \end{cases}$$
then
$$L\{G(t)\} = \exp(-as) \, F(s). \qquad (2.4)$$

4. *Change of scale property*:
$$L\{f(at)\} = (1/a) \, F(s/a). \qquad (2.5)$$

Since $L\{\sin t\} = \dfrac{1}{s^2+1}$, we have
$$L\{\sin at\} = (1/a) \frac{1}{(s/a)^2 + 1} = \frac{a}{s^2 + a^2}.$$

Example 2(b).

$$L\{(\exp(-at))\,(at)^n\} = L\{f(at)\} = (1/a)\,F(s/a)$$

$$= \frac{\Gamma(n+1)}{a((s/a)+1)^{n+1}} = \frac{a^n\,\Gamma(n+1)}{(s+a)^{n+1}}.$$

5. *L.T. of derivatives*: $L\{f'(t)\} = sF(s) - f(0)$. $\quad$ (2.6)

Example 2(c). If $f(t) = \sin t$, then $F(s) = 1/(s^2+1)$

$$L\{\cos t\} = L\{f'(t)\} = s/(s^2+1).$$

6. *L.T. of integrals* : $L\left\{\int\limits_0^t f(x)\,dx\right\} = F(s)/s.$ $\quad$ (2.7)

If $\qquad f(t) = \cos t$, then $F(s) = s/(s^2+1)$.

Since $\qquad \sin t = \int\limits_0^t \cos t\,dt$, $\;L\{\sin t\} = 1/(s^2+1)$.

7. *Limit property*: Initial value theorem

$$\lim_{t\to\infty} f(t) = \lim_{s\to 0} sF(s), \tag{2.8}$$

whenever the limits exist.

Example 2(d). Let $f(t) = \exp(-t)$, then $F(s) = 1/(s+1)$.
We have $\lim\limits_{t\to\infty} f(t) = 0$, as is $\lim\limits_{s\to 0} sF(s)$.

8. *Limit property:* Final value theorem

$$\lim_{t\to 0} f(t) = \lim_{s\to\infty} sF(s), \tag{2.9}$$

whenever the limits exist.

Example 2(e). Let $f(t) = \exp(-t)$, then $F(s) = 1/(s+1)$

$$\lim_{t\to 0} f(t) = 1 \text{ and } \lim_{s\to\infty} sF(s) = \lim_{s\to\infty} \frac{s}{s+1} = 1$$

The properties (7) and (8) enable one to find the behaviour of $f(t)$ as $t \to \infty$ (or as $t \to 0$) from the behaviour of $F(s)$ as $s \to 0$ (or as $s \to \infty$). There are many other such results of great generality; they are covered under the name of Tauberian theorems.

C. Inverse Laplace Transform

Definition. If $F(s)$ is the L.T. of $f(t)$, i.e. $L\{f(t)\} = F(s)$, then $f(t)$ is called the inverse Laplace transform of $F(s)$. For example, the inverse Laplace transform of $F(s) = 1/s^2$ is $f(t) = t$. There is an inversion formula which gives $f(t)$ in terms of $F(s)$ and given $F(s)$, the formula could be employed to find $f(t)$ (uniquely).

Extensive tables of Laplace transforms are also available and reference may be made to them, whenever necessary, to obtain either the L.T. of a function $f(t)$ or the inverse L.T. of a function $F(s)$.

1.3 Laplace (Stieltjes) Transform of a Probability Distribution or of a Random Variable

Let X be an non-negative random variable with distribution function

$$F(x) = \Pr\{X \leqslant x\}.$$

The Laplace (Laplace-Stieltjes) transform $f^*(s)$ of this distribution is defined, for $s \geqslant 0$, by

$$f^*(s) = \int_0^\infty \exp(-sx)\, dF(x) \tag{3.1}$$

(assuming that the range of X is from 0 to ∞).

We shall say that (3.1) also gives the "Laplace (Laplace-Stieltjes) transform of the random variable X".

We have

$$f^*(s) = E\{\exp(-sX)\} \tag{3.1a}$$

and

$$f^*(0) = 1. \tag{3.2}$$

Suppose that X is a continuous variate having density $f(x) = F'(x)$. Then from (3.1),

$$f^*(s) = \int_0^\infty \exp(-sx) f(x) dx \tag{3.3}$$

(this is the "ordinary" Laplace transform $L\{f(x)\}$ of the density function $f(x)$). In case X is an integral-valued random variable with distribution $p_k = \Pr\{X = k\}$, $k = 0, 1, 2, \ldots$, and p.g.f. $P(s) = \sum_k p_k s^k$, we can stretch the language and define the L.T. of X by

$$f^*(s) = E\{\exp(-sX)\} = P\{\exp(-s)\}. \tag{3.4}$$

Thus in case of a discrete random variable assuming integral values $0, 1, 2, 3, \ldots$, the L.T. of the variable differs from its p.g.f. only by a change of variable: $\exp(-s)$ in the former replaces s in the latter. Thus there is a close analogy between the properties of Laplace transforms and those of generating functions or s-transforms.

The Laplace Transform of the Distribution Function in Terms of that of the Density Function

Let X be a continuous (and non-negative) r.v. having density function $f(x)$ and distribution function

$$\Pr\{X \leqslant x\} = F(x) = \int_0^x f(t)\, dt$$

The (ordinary) Laplace transform of the distribution function $F(x)$ is

$$L\{F(x)\} = \int_0^\infty \exp(-sx) F(x) dx$$

and the Laplace transform (Laplace-Stieltjes transform) of the distribution (or of the r.v. X) is

$$f^*(s) = \int_0^\infty \exp(-sx) dF(x) = \int_0^\infty \exp(-sx) f(x) dx = L\{f(x)\}.$$

We have from (2.7)

$$L\{F(x)\} = \int_0^\infty \exp(-sx) \left\{\int_0^x f(t)dt\right\} dx = L\left\{\int_0^x f(t)dt\right\}$$

$$= f^*(s)/s. \tag{3.5}$$

The relation can also be obtained by integrating by parts the relation (3.1).

A. Mean and Variance in Terms of (Derivatives of) L.T.

We note here that differentiation under the integral sign is valid for the L.T. given by (3.1).

Differentiating (3.1) with respect to s, we get

$$\frac{d}{ds} f^*(s) = \int_0^\infty (-x) \exp(-sx) dF(x)$$

$$= -\int_0^\infty x \exp(-sx) dF(x),$$

$$\frac{d^2}{ds^2} f^*(s) = (-1)^2 \int_0^\infty x^2 \exp(-sx) dF(x)$$

and, in general,

$$\frac{d^n}{ds^n} f^*(s) = (-1)^n \int_0^\infty x^n \exp(-sx) dF(x). \tag{3.6}$$

We can use the above relation to find $E(X^n)$, the nth moment of X; we have

$$(-1)^n \left[\frac{d^n}{ds^n} f^*(s)\right]_{s=0} = \int_0^\infty \{\exp(-sx)\}\, x^n\, dF(x) \text{ for } s=0$$

$$= \int_0^\infty x^n\, dF(x) = E(X^n).$$

In particular,

$$E(X) = -\left[\frac{d}{ds} f^*(s)\right]_{s=0} \tag{3.7}$$

$$E(X^2) = \left[\frac{d^2}{ds^2} f^*(s)\right]_{s=0} \tag{3.8}$$

and

$$\operatorname{var}(X) = \left[\frac{d^2}{ds^2} f^*(s)\right]_{s=0} - \left[\left\{-\frac{d}{ds} f^*(s)\right\}_{s=0}\right]^2. \tag{3.9}$$

Example 3(a). A special kind of discrete distribution:
Suppose that X is a random variable whose whole mass is concentrated in one single point, say, point a. This implies that the variable is 'almost always' equal to a. i.e. $\Pr(X=a)=1$ and $\Pr(X \neq a) = 0$. Its distribution function is given by

$$\begin{aligned} F(x) = \Pr\{X \leqslant x\} &= 0 \qquad \text{for} \quad x < a \\ &= 1 \qquad \text{for} \quad x \geqslant a. \end{aligned}$$

We have $E(X) = a$ and $\operatorname{var}(X) = E(X^2) - (E(X))^2 = a^2 - a^2 = 0$.

The L.T. of the distribution is given by

$$f^*(s) = E\{\exp(-sX)\} = \exp(-sa)\Pr(X=a) = \exp(-sa). \tag{3.10}$$

This is the L.T. of Dirac function located at a.

Example 3(b). *Poisson distribution:* Let X have Poisson distribution with mean λ. Its p.g.f. is $P(s) = \exp\{\lambda(s-1)\}$.

From (3.4), its L.T. is given by

$$f^*(s) = P\{\exp(-s)\} = \exp[\lambda\{\exp(-s) - 1\}]. \tag{3.11}$$

Example 3(c). *Negative exponential distribution:* Let X have the negative exponential distribution with parameter λ. Its density function is

$$\begin{aligned} f(x) &= \lambda \exp(-\lambda x), \qquad \lambda > 0, \qquad 0 \leqslant x < \infty, \\ &= 0 \qquad\qquad\qquad\qquad\quad , \qquad x < 0. \end{aligned} \tag{3.12}$$

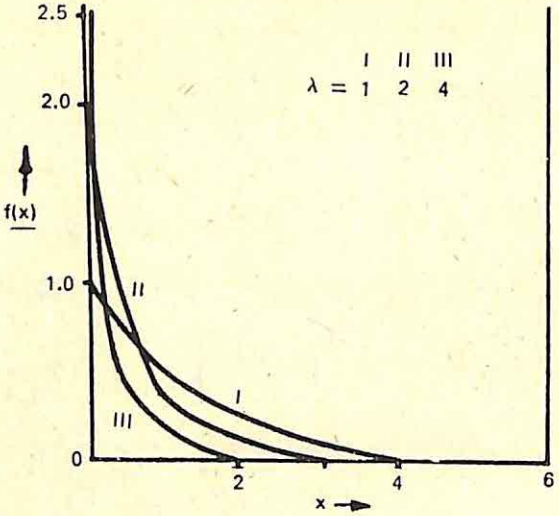


Fig. 1.1. The negative exponential density function for some values of λ.

The L.T. $f^*(s)$ of the r.v. X is

$$f^*(s) = \int_0^\infty \{\exp(-sx)\}\{\lambda \exp(-\lambda x)\}\, dx = \lambda/(s+\lambda).$$

We have

$$\frac{d}{ds} f^*(s) = -\frac{\lambda}{(s+\lambda)^2}, \quad \frac{d^2}{ds^2} f^*(s) = \frac{2\lambda}{(s+\lambda)^3}.$$

Hence from (3.7) and (3.8)

$$E(X) = -\left[-\frac{\lambda}{\lambda^2}\right] = \frac{1}{\lambda}, \quad E(X^2) = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2} \tag{3.13}$$

and $$\operatorname{var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \tag{3.14}$$

The distribution is also known as simply *exponential distribution.*

Example 3(d). *Uniform (rectangular) distribution:* Let X have uniform distribution in (0, 1). The density function of X is

$$\begin{aligned} f(x) &= 1, && 0 \leqslant x \leqslant 1 \\ &= 0, && \text{otherwise.} \end{aligned}$$

The L.T. $f^*(s)$ of the r.v. X is

$$f^*(s) = \int_0^\infty \exp(-sx)\, f(x)\, dx = \int_0^1 \exp(-sx)\, dx = \{1 - \exp(-s)\}/s.$$

The mean of X is $\frac{1}{2}$ and the variance is $1/12$.

Example 3(e). *Gamma distribution:* Let X have a two parameter gamma distribution with parameters λ, k ($\lambda > 0$ is the *scale* parameter and $k > 0$ is the *shape* parameter).

The density function of the r.v. X is

$$f_{\lambda,k}(x) = \frac{\lambda^k x^{k-1} \exp(-\lambda x)}{\Gamma(k)}, \quad x > 0$$
$$= 0, \quad x < 0. \tag{3.15}$$

The L.T. $f^*(s)$ is given by

$$f^*(s) = \int_0^\infty \exp(-sx)\, f_{\lambda,k}(x)\, dx$$

$$= (\lambda^k/\Gamma(k)) \int_0^\infty [\exp\{-(s+\lambda)x\}]\, x^{k-1}\, dx$$

$$= (\lambda^k/\Gamma(k)) \int_0^\infty \frac{\exp(-t)\, t^{k-1}}{(s+\lambda)^k}\, dt, \text{ (putting } (s+\lambda)\, x = t)$$

$$= \left(\frac{\lambda}{s+\lambda}\right)^k. \tag{3.16}$$

We have

$$\frac{d}{ds} f^*(s) = -\frac{k\lambda^k}{(s+\lambda)^{k+1}},$$
$$\frac{d^2}{ds^2} f^*(s) = \frac{k(k+1)\,\lambda^k}{(s+\lambda)^{k+2}}.$$

Hence from (3.7) and (3.8)

$$E(X) = -\left[-\frac{k\lambda^k}{\lambda^{k+1}}\right] = \frac{k}{\lambda}, \tag{3.17}$$

$$E(X^2) = \frac{k(k+1)\,\lambda^k}{\lambda^{k+2}} = \frac{k(k+1)}{\lambda^2},$$

and

$$\text{var}\,(X) = \frac{k(k+1)}{\lambda^2} - \left(\frac{k}{\lambda}\right)^2 = \frac{k}{\lambda^2}. \tag{3.18}$$

When $\lambda = 1$, we get single parameter gamma variate with density

$$\frac{x^{k-1} \exp(-x)}{\Gamma(k)} \; (x > 0);$$

and its Laplace transform is $1/(s+1)^k$.

When $k = 1$, the gamma distribution becomes the negative exponential distribution with density $\lambda \exp(-\lambda x)$ and having Laplace transform $\lambda/(s+\lambda)$. Another special case is $\lambda = 1/2$ and $k = n/2$, where n is a positive integer. The density then becomes (for $x > 0$)

$$\frac{x^{n/2-1} \exp(-x/2)}{2^{n/2}\,\Gamma(n/2)}$$

and the random variable is then said to have a *chi-square* (χ^2) distribution with n degrees of freedom. The L.T. of the density is $(1+2s)^{-n/2}$.

Figure 1.2 gives the density function of gamma distributions for some values of λ and k.

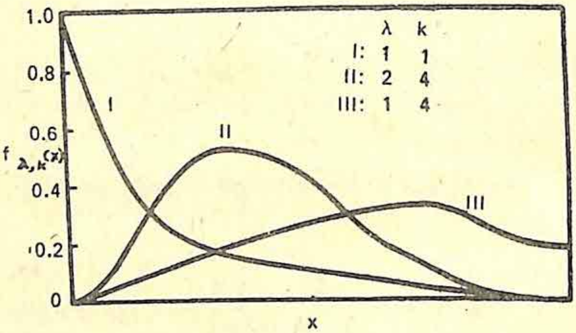


Fig. 1.2. The gamma density $f_{\lambda,k}(x)$ for some values of λ, k.

The gamma distribution has considerable scope for practical application; many empirical distributions can be represented roughly by a suitable choice of the parameters λ, k.

Example 3(f). *Erlang distribution*: Writing λk for λ, we have a gamma distribution with parameters λk, k having density function

$$f_{\lambda k,\,k} = \frac{(\lambda k)^k x^{k-1} \exp(-k\lambda x)}{\Gamma(k)}, \quad x > 0 \tag{3.19}$$

The variable having this for its density function is said to have *Erlang* (or *Erlang—k*) distribution and is denoted by E_k. In fact, when k is a positive integer, gamma distribution with density $f_{\lambda k,k}$ is known as *Erlang* distribution.

The L.T. of *Erlang* distribution is

$$\left(\frac{\lambda k}{s+\lambda k}\right)^k ; \tag{3.20}$$

further, $E(E_k) = k/\lambda k = 1/\lambda$; $\operatorname{var}(E_k) = k/(\lambda k)^2 = 1/k\lambda^2$. (3.21)

B. Three Important Theorems

We state without proof three theorems concerning L.T. of probability distributions.

Theorem 1.5. *Uniqueness Theorem:* Distinct probability distributions have distinct Laplace transforms.

It has the important implication that a probability distribution is recognizable by its transform. If the L.T. of a random variable X is known, then by identifying this with the form of L.T. of a distribution, one can conclude that X has the corresponding distribution; if this form does not resemble with the form of a L.T. of any standard distribution one can proceed to find the inverse of the L.T. (some numerical methods have also been developed) to get the distribution. Even without finding the inverse, i.e. the form of the distribution, one can compute the moments of the distribution. But for this theorem, perhaps, L.T. would not have been able to play the role it now plays in the study of probability distributions.

Theorem 1.6. *Continuity Theorem*: Let $\{X_n\}$, $n = 1, 2, \ldots$ be a sequence of random variables with distribution function $\{F_n\}$ and L.T. $\{f_n^*(s)\}$. If as $n \to \infty$, F_n tends to a distribution function F having transform $f^*(s)$, then as $n \to \infty$, $f_n^*(s) \to f(s)$ for $s > 0$, and conversely.

This theorem can be used to obtain the limit distribution of a sequence of random variables.

Theorem 1.7. *Convolution theorem*: The Laplace transform of the convolution of two independent random variables X, Y is the product of their transforms.

The integral

$$\int_0^x g(x-y)\, f(y)\, dy$$

(denoted by f^*g) is called the convolution of the two functions f and g. The convolution U of F and G is given by

$$U(x) = \int_0^x G(x-y)\, dF(y) \tag{3.22}$$

and

$$L\{U(x)\} = L\{G(x)\}\, L\{F(x)\}. \tag{3.22a}$$

In case of discrete random variables the result is essentially the same as that stated in (1.10) for generating function of the convolution of two random variables. In case of continuous random variables we have now the analogous result.

The above result is equivalent to the assertion that if X, Y are two independent random variables, then

$$\begin{aligned} E\{e^{-s(X+Y)}\} &= E\{\exp(-sX)\exp(-sY)\} \\ &= E\{\exp(-sX)\}\, E\{\exp(-sY)\}. \end{aligned} \tag{3.23}$$

The result can also be stated as:

519.2
MED
S.C.E.R.T., West Bengal
10-3-87
S.C.E.R.T., West Bengal Library

Theorem 1.8. The L.T. of the sum of two independent variables is the product of the L.T.'s of the variables.

We can use the result to find the distribution of the sum of two or more independent random variables, when the variables are continuous or discrete. The method of generating functions is applicable only when the variables are non-negative integral-valued.

The following result immediately follows:

Theorem 1.9. The sum $S_n = X_1 + \ldots + X_n$ of n (a *fixed* number) identically and independently distributed random variables X_i has the L.T. equal to $[f^*(s)]^n$. $f^*(s)$ being the L.T. of X.

As an application we consider the following result:

Theorem 1.10. The sum of k identical and independent negative exponential distributions with parameter λ follows gamma distribution with parameters λ, k.

Let $\quad X_i,\ i = 1, 2, \ldots, 1, 2, \ldots, k,\quad$ have densities

$$f_i(x) = \lambda \exp(-\lambda x),\ x \geqslant 0 \quad \text{for all } i.$$

From Example 3(c), the L.T. of X_i is

$$f_i^*(s) = \lambda/(s+\lambda)$$

and so the L.T. of $S = X_1 + X_2 + \ldots + X_k$ is $(\lambda/(s+\lambda))^k$.
But this is the L.T. of the gamma distribution with density $f_{\lambda,k}(x)$ given in (3.15) (Example 3(c)). Hence the theorem. ▲

We can easily obtain the mean and variance of the gamma distribution as follows:

$$E(S) = E(\Sigma X_i) = \Sigma E(X_i) = k/\lambda$$

$$\text{var}(S) = \Sigma \text{ var}(X_i) = k/\lambda^2 \quad \text{(see (3.17) and (3.18))}.$$

We shall show that there is a close analogy between the (discrete) geometric distribution and the (continuous) negative exponential distribution (Sec. 4.4). Negative binomial distribution, which can be obtained as the sum of a number of independent geometric distributions (Example 1(g)) has the continuous analogue in the gamma distribution which can be obtained as the sum of independent exponential distributions.

Limiting form of Erlang distribution: As $k \to \infty$, *Erlang* distribution tends to a discrete distribution concentrated at $1/\lambda$. Let $\{E_k\}$ be a sequence of r.v.'s, E_k having *Erlang* distribution with density

$$f_{\lambda k,\,k} = \frac{(\lambda k)^k\, x^{k-1} \exp(-k\lambda x)}{\Gamma(k)},\ k = 1.2.3, \ldots.$$

Its L.T. is given by $f_k^*(s) = \left(\dfrac{\lambda k}{s+\lambda k}\right)^k = \left(1 + \dfrac{s}{\lambda k}\right)^{-k}$.

As $$k \to \infty,\ f_k^*(s) \to \exp(-s/\lambda). \tag{3.24}$$

But $\exp(-s/\lambda)$ is the L.T. of the discrete distribution concentrated at $a = 1/\lambda$ (Example 3(a)). Therefore, from the continuity theorem it follows that, as $k \to \infty$, the distribution of E_k tends to that variable whose whole mass is concentrated at $1/\lambda$.

This can also be seen from the fact that, as $k \to \infty$, the mode $(k-1)/k\lambda$ of E_k moves to the right to $1/\lambda$ and that as $\text{var}(E_k) = 1/k\lambda^2 \to 0$, the whole mass of E_k tends to concentrate at the single point $1/\lambda$.

We thus have exponential distribution for $k = 1$ and the deterministic case for $k \to \infty$. A suitable value of k $(1 < k < \infty)$ gives a good fit for many empirical distributions arising in practice.

An idea of how the density varies as k increases can be seen from Fig. 1.3.

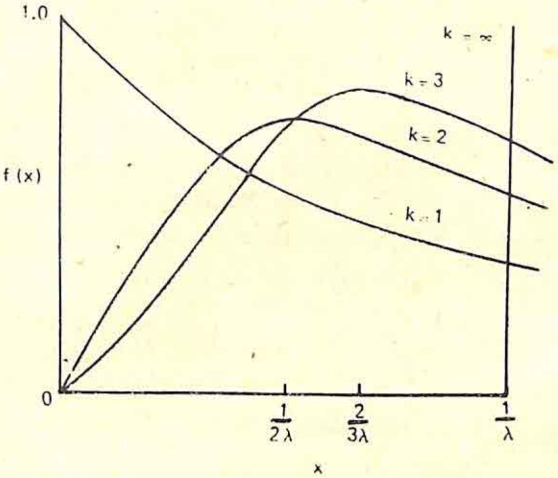


Fig. 1.3. The density function of *Erlang* distribution for some values of k

C. Sum of a Random Number of Continuous Random Variables

In Sec. 1.1. we considered the sum S_N of a random number N of random variables X_i. Theorem 1.3 is stated in terms of the p.g.f. of X_i on the assumption that the X_i's are discrete variables (with non-negative integral values). Here we assume that the X_i's are mutually independent and identically distributed variables which may be continuous as well.

Replacing the p.g.f. $P(s)$ of X_i by its L.T. $f^*(s)$, we can get an analogous result. This may be stated as follows:

Theorem 1.11. If X_i are identically and independently distributed random variables having for their L.T.'s

$$f^*(s) = E\{\exp(-sX_i)\} \qquad \text{for all } i \tag{3.25}$$

and if N (independent of X_i's) is a random variable having p.g.f.

$$G(s) = \sum_n \Pr\{N = n\}\, s^n = \sum_n g_n s^n,$$

then the L.T. $F(s)$ of the sum $S_N = X_1 + \ldots + X_N$,

is given by
$$F(s) = G(f^*(s)). \tag{3.26}$$

The proof is identical with that of Theorem 1.3. It also follows that
$$E\{S_N\} = E\{X_i\}\, E\{N\}. \tag{3.27}$$

It is to be noted that S_N is a continuous random variable.

Corresponding to $h_j = \Pr\{S_N = j\}$ we shall have density function $f(x)$ of S_N, so that

$$f(x)\, dx = \Pr\{x \leqslant S_N < x + dx\} = \sum_{n=0}^{\infty} \Pr\{N = n \text{ and } x \leqslant S_n < x + dx\} \tag{3.28}$$

$$= \sum_{n=0}^{\infty} \Pr\{N = n\} \Pr\{x \leqslant S_n < x + dx\}$$
$$= \sum_n g_n \Pr\{x \leqslant S_n < x + dx\}. \tag{3.29}$$

Example 3(g). *Random (geometric) sum of exponential variables:*
If X_i, $i = 1, 2, \ldots$ have negative exponential distribution each with mean $1/\lambda$ and $S_N = X_1 + \ldots + X_N$, where N has geometric distribution given by $\Pr\{N = n\} = pq^{n-1}$, $n = 1, 2, \ldots$ then S_N has negative exponential distribution with mean $1/\lambda p$.

The p.g.f. of N is $G(s) = ps/(1 - qs)$

and the L.T. of the r.v. X_i is $f^*(s) = \lambda/(s + \lambda)$.

From the Theorem 1.11, it follows that the sum $S_N = X_1 + \ldots + X_N$ has L.T. equal to $G(f^*(s))$. Replacing s in $G(s)$ by $f^*(s) = \lambda/(s + \lambda)$, we have

$$G(f^*(s)) = \frac{p\lambda/(s + \lambda)}{1 - q\lambda/(s + \lambda)} = \frac{p\lambda}{s + p\lambda}; \tag{3.30}$$

but this is the L.T. of the exponential distribution with mean $1/\lambda p$. Hence the probability element $f(x)\, dx$ of S_N is given by

$$f(x)\, dx = \Pr\{x \leqslant S_N < x + dx\} = \lambda p \exp(-\lambda p x)\, dx. \tag{3.31}$$

Example 3(h). Let X_i have exponential distribution with mean $1/a$ and let N have Poisson distribution with mean c. We have $f^*(s) \equiv$ L.T. of $X_i = a/(s + a)$, and $G(s) = e^{c(s-1)}$ and hence the L.T. of $S_N = X_1 + \ldots + X_N$ is $G(f^*(s))$

$$= \exp\left(c\left(\frac{a}{s + a} - 1\right)\right), \text{ replacing } s \text{ by } f^*(s) = a/(s + a)$$
$$= \exp(-cs/(s + a)).$$

What is the probability density function of S_N?

1.4 Difference Equations

Let $f(n)$ be a function defined only for non-negative *integral values* of the argument n. The first difference of $f(n)$ is defined by the increment of $f(n)$ and is denoted by $\Delta f(n)$, i.e.

$$\Delta f(n) = f(n+1) - f(n). \tag{4.1}$$

The second and the higher differences are defined by

$$\Delta^{k+1} f(n) = \Delta\{\Delta^k f(n)\}, \quad k > 0.$$

It can be verified that

$$\begin{aligned} \Delta^{k+1} f(n) &= \Delta^k\{\Delta f(n)\} = \Delta^k\{f(n+1) - f(n)\} \\ &= \Delta^k f(n+1) - \Delta^k f(n),\ k > 0. \end{aligned} \tag{4.2}$$

We denote $f(n+1)$ by $Ef(n)$ (E is here the displacement operator and is not to be confused with the symbol E for the expectation of a random variable): then

$$\Delta f(n) = Ef(n) - f(n) = (E-1)f(n).$$

Symbolically, $\Delta \equiv E - 1$ or $E \equiv 1 + \Delta$; we have

$$E^k f(n) = f(n+k). \tag{4.3}$$

By a difference equation, we mean an equation involving a function evaluated at arguments which differ by any of a fixed number of values. Examples of such equations are:

$$\begin{aligned} &f(n+2) - f(n+1) - f(n) = 0 \\ &(\lambda + \mu_n)p_n = \mu_{n+1} p_{n+1} + \lambda p_{n-1},\ n \geqslant 1 \\ &p_n - q p_{n-1} - p(1 - p_{n-1}) = 0 \\ &Y_t = a(1+b)Y_{t-1} - abY_{t-2} + 1 \\ &a_0 U_x + a_1 U_{x+1} + \ldots + a_k U_{x+k} = g(x). \end{aligned} \tag{4.4}$$

In the first equation the argument runs from n to $(n+2)$; this is a difference equation of order $(n+2) - n = 2$. The last equation is of order $(x+k) - x = k$. The first two equations are homogeneous while the last three are not.

The equation $f(n+2) - f(n+1) - f(n) = 0$ may be written as

$$E^2 f(n) - Ef(n) - f(n) = 0, \text{ or } r(E)f(n) = 0,$$

where $r(E) = (E^2 - E - 1)$.

A difference equation may be written as a function of the operator E in the same way that a differential equation can be written as a function of the operator $D\left(\equiv \dfrac{d}{dx}\right)$. There is indeed a close similarity between the theory of differential equations and that of difference equations.

Given a function $f(n) = Ca^n$, we can easily get the difference equation $f(n+1) - af(n) = 0$. We shall be generally concerned with the inverse problem: given a difference equation, to find the function which satisfies the given difference equation.

GENERAL SOLUTION OF A HOMOGENEOUS DIFFERENCE EQUATION WITH CONSTANT COEFFICIENTS

Consider a homogeneous difference equation of order *two*

$$f(n+2)+af(n+1)+bf(n)=0 \tag{4.5}$$

or $$r(E)f(n)=0, \text{ where } r(E)=E^2+aE+b.$$

Consider a trial solution $f(n)=m^n \quad (m \neq 0)$;

then $Ef(n)=m^{n+1}$, $E^2f(n)=m^{n+2}$ so that (4.5) gives

$$(m^2+am+b)=0. \tag{4.6}$$

This equation $r(m)=0$ is called the *characteristic* (or auxiliary or secular) *equation* of the difference equation (4.5).

(i) If the roots m_1, m_2 of the equation (4.6) are distinct, then $m_1{}^n$ and $m_2{}^n$ are two solutions of the equation (4.5). Since the equation is homogeneous $c_1 m_1{}^n$ and $c_2 m_2{}^n$ will also satisfy the equation and so also will their sum. In other words, if m_1, m_2 are the distinct roots of (4.6) and c_1, c_2 are arbitrary constants, then

$$f(n)=c_1m_1{}^n+c_2m_2{}^n \tag{4.7}$$

will be the general solution of the equation (4.5).

(ii) In case the roots m_1, m_2 are equal, the general solution is given by

$$f(n)=(c_1+c_2n)m_1{}^n. \tag{4.8}$$

(iii) In case the roots are complex, then we can write

$$m_1=r(\cos\theta+i\sin\theta) \text{ and } m_2=r(\cos\theta-i\sin\theta)$$

and the general solution (4.7) will take the form

$$f(n)=Ar^n\cos(n\theta+B), \tag{4.9}$$

where A, B are arbitrary constants.

The above arguments can be carried over to homogeneous difference equations of any order k. We may state as follows the rule for the solution of the homogeneous difference equations:

$$f(n)+a_1f(n+1)+\ldots+a_kf(n+k)=0 \tag{4.10}$$

or, $$r(E)f(n)=0.$$

First, find the roots of the characteristic equation

$$r(m)=1+a_1m+\ldots+a_km^k=0. \tag{4.11}$$

Suppose that $m_1, m_2, \ldots, m_k$ are the real and distinct roots of (4.11), then the general solution of (4.10) will be given by

$$f(n)=c_1m_1{}^n+c_2m_2{}^n+\ldots+c_km_k{}^n. \tag{4.12}$$

In case some of the real roots are equal, say $m_1=m_2$, and others distinct then the solution will be

$$f(n)=(c_1+c_2n)m_1{}^n+c_3m_3{}^n+\ldots+c_km_k{}^n. \tag{4.13}$$

It is to be noted that the general solution of an equation of order k will, in general, contain k arbitrary constants $c_1, c_2, \ldots, c_k$. These can be obtained when k initial (or boundary) conditions are also given.

General solution of a general linear equation $r(E)\,f(n) = g(n)$ is given by $u + v$, where u is a solution of the homogeneous equation $r(E)f(n) = 0$ and v is a particular solution of the complete equation. Methods of obtaining this particular solution depend largely on the form of the function $g(n)$.

There are other methods also for solving a homogeneous difference equation. Some of these, which could be used to solve certain types of non-homogeneous equations, are discussed in the next section.

Example 4(a). *Fibonacci* numbers are given by

$$f_0 = 0,\ f_1 = 1,\ f_n = f_{n-1} + f_{n-2} \quad (n \geqslant 2).$$

The difference equation $f_n = f_{n-1} + f_{n-2}$ has the characteristic equation $m^2 = m + 1$, whose roots are $m_1 = (1 + \sqrt{5})/2$ and $m_2 = (1 - \sqrt{5})/2$. Hence $f_n = c_1 m_1{}^n + c_2 m_2{}^n$. Solving for c_1 and c_2 by using $f_0 = 0, f_1 = 1$, we get $c_1 = -c_2 = 1/\sqrt{5}$, so that

$$f_n = [\{(1+\sqrt{5})/2\}^n - \{(1 - \sqrt{5})/2\}^n]/\sqrt{5}.$$

1.5 Difference Equations in Probability Theory

Many problems in probability theory can be put down in terms of a difference equation involving the unknown probability. Such problems can then be solved completely by making use of the initial conditions supplied by the problem itself. This method is sometimes found to be very powerful.

Example 5(a). To solve the difference equation

$$U_n = \tfrac{1}{2}(U_{n+1} + U_{n-1}),\ 1 \leqslant n \leqslant a - 1$$

having initial conditions

$$U_0 = 1 \text{ and } U_a = 0.$$

The difference equation has the characteristic equation $\frac{1}{2}m^2 - m + \frac{1}{2} = 0$ having two equal roots $m = 1, 1$. The solution of the difference equation is

$$U_n = (c_1 + c_2 n)(1)^n = c_1 + c_2 n.$$

The initial conditions give

$$1 = U_0 = c_1,\ 0 = U_a = c_1 + c_2 a, \quad \text{whence } c_2 = -1/a.$$

The solution of the difference equation is thus

$$U_n = 1 + (-1/a)n = 1 - n/a.$$

(The equation arises in the classical ruin problem, U_n being the probability of gambler's ruin in the symmetric case $p = q = \frac{1}{2}$.) In the study of stochastic processes, we frequently come across equations which are non-homogeneous and which have coefficients that are not constants but functions of the argument. The method of induction and the method of generating function could be used in such cases. These methods are illustrated by an example.

Example 5(b). The probability function

$$p_n = \Pr\{N = n\}, \qquad n = 0, 1, 2, \ldots,$$

of a random variable N satisfies the difference equations:

$$p_{n+1} - (1 + a)\, p_n + ap_{n-1} = 0,\ n \geqslant 1 \tag{5.1}$$

$$-p_1 + ap_0 = 0 \qquad (0 < a < 1). \tag{5.2}$$

Since p_n's are probabilities of the random variable N, we have also the condition

$$\sum_{n=0}^{\infty} p_n = 1. \tag{5 3}$$

We now proceed to find p_n, *i.e.* to solve the difference equation (5.1) given (5.2) and (5.3).

METHOD OF CHARACTERISTIC EQUATION

The second order difference equation (5.1) which can also be written as

$$p_{n+2} - (1 + a)\, p_{n+1} + ap_n = 0,\ n \geqslant 0 \tag{5.4}$$

has the characteristic equation $m^2 - (1 + a)\, m + a = 0$, whose roots are $m = 1$ and $m = a$ (< 1). So the equation (5.1) has the solution

$$p_n = A\,(1)^n + B(a)^n = A + Ba^n,\ n \geqslant 0. \tag{5.5}$$

We have from (5.4) and (5.2)

$$p_1 = A + Ba = ap_0$$

$$p_0 = A + B. \tag{5.6}$$

Solving in terms of p_0, we get $A = 0$, $B = p_0$, so that

$$p_n = p_0 a^n. \tag{5.7}$$

From (5.3) we get

$$1 = \Sigma\, p_n = \Sigma\, p_0 a^n = p_0/(1 - a) \quad \text{or} \quad p_0 = 1 - a.$$

Hence the solution is

$$p_n = (1 - a)\, a^n,\ n \geqslant 0. \tag{5.8}$$

METHOD OF INDUCTION

The equation $p_{n+1} - (1 + a)\, p_n + ap_{n-1} = 0$ can be written as

$$p_{n+1} - ap_n = p_n - ap_{n-1},\ n \geqslant 1.$$

Putting $(n-1)$ for n, we get

$$p_n - ap_{n-1} = p_{n-1} - ap_{n-2}$$

and so on. Hence

$$p_{n+1} - ap_n = p_n - ap_{n-1} = p_{n-1} - ap_{n-2} = \dots$$
$$= p_1 - ap_0, \qquad n \geqslant 0. \qquad (5.9)$$

But from (5.2) $p_1 - ap_0 = 0.$

Hence

$$p_n = ap_{n-1} = a(ap_{n-2}) = \dots = a^n p_0,$$

and

$$\sum_{n=0}^{\infty} p_n = 1 \text{ gives } p_0 = 1 - a, \text{ so that } p_n = (1-a)\, a^n, \; n \geqslant 0.$$

METHOD OF GENERATING FUNCTIONS

Let

$$P(s) = \sum_{n=0}^{\infty} p_n s^n \qquad (5.10)$$

be the p.g.f. of the distribution.

Multiplying equation (5.1) by $s^n (n \geqslant 1)$ and adding over $n = 1, 2, \dots$, we get

$$\sum_{n=1}^{\infty} p_{n+1}\, s^n - (1+a) \sum_{n=1}^{\infty} p_n\, s^n + a \sum_{n=1}^{\infty} p_{n-1}\, s^n = 0. \qquad (5.11)$$

or, $(1/s)\,[P(s) - p_0 - p_1 s] - (1+a)\,[P(s) - p_0] + asP(s) = 0$

or, $[1/s - (1+a) + as]\; P(s) = \dfrac{p_0}{s} + p_1 - (1+a)\, p_0$

or, $\dfrac{(1-s)(1-as)}{s}\; P(s) = \dfrac{p_0\,(1-s)}{s}.$

Hence

$$P(s) = \frac{p_0}{1-as} = p_0\,(1-as)^{-1} = p_0\,(1 + as + a^2 s^2 + \dots)$$

so that

$$p_n \equiv \text{coefficient of } s^n \text{ in } P(s) \qquad (5.12)$$
$$= p_0 a^n.$$

Finding p_0 as before from (5.3), we get

$$p_n = (1-a)\, a^n, \quad n \geqslant 0.$$

The method of Laplace transforms can also be used to solve a difference equation (Smith Ch. 9).

Example 5(c). To solve the difference equation

$$u_n = qu_{n-1} + p(1 - u_{n-1}), \; n \geqslant 1, \; p + q = 1, \qquad (5.13)$$

given the initial condition $u_0 = 1$. It is a nonhomogeneous equation of order 1. Only one initial condition is needed to solve it completely.

Let $A(s) = \sum_{n=0}^{\infty} u_n s^n$ be the generating function of $\{u_n\}$ given by (5.13). Multiplying (5.13) by s^n and adding over $n = 1, 2, \ldots$, we get

$$\sum_{n=1}^{\infty} u_n s^n = (q-p) \sum_{n=1}^{\infty} u_{n-1} s^n + p \sum_{n=1}^{\infty} s^n$$

or, $$(A(s) - u_0) = (q-p) sA(s) + p \cdot \frac{s}{1-s}$$

or, $$\{1-(q-p)s\} A(s) = \frac{ps}{(1-s)} + 1 \quad (\text{since } u_0 = 1) \tag{5.14}$$

or, $$A(s) = \frac{ps}{(1-s)\{1-(q-p)s\}} + \frac{1}{1-(q-p)s}$$

$$= \tfrac{1}{2}\left\{\frac{1}{1-s} - \frac{1}{1-(q-p)s}\right\} + \frac{1}{1-(q-p)s},$$

(decomposing into partial fractions)

$$= \tfrac{1}{2}\{(1-s)^{-1} + [1-(q-p)s]^{-1}\}.$$

Thus

$$u_n \equiv \text{coefficient of } s^n \text{ in } A(s) = \tfrac{1}{2}\{1+(q-p)^n\}.$$

(Here u_n gives the probability that n Bernoulli trials result in an even number of successes. $\{u_n\}$ is not a probability distribution.)

The method of generating functions is very powerful. It can also be used when the coefficients of the equation are not constants, but are functions of the argument.

1.6 Differential-Difference Equations

Suppose that $u_n(t)$, $n = 1, 2, \ldots$, is a function of t having a derivative

$$\frac{d}{dt} u_n(t) = u_n'(t).$$

An equation involving $u_n'(t)$ and $u_n(t)$, $u_{n+1}(t)$ etc. is called a differential-difference (or recurrence differential) equation. For example,

$$p_n'(t) = -\lambda [p_n(t) - p_{n-1}(t)]$$

is a differential-difference equation (see Bellman and Cooke for a detailed account of this topic).

The existence of unique solution will depend on suitable initial conditions which are to be determined carefully. The solution of such equations can sometimes be found easily by the method of induction (see Example 5(b)). Besides, there are two powerful tools for the solution of such an equation: these are generating functions and Laplace transforms. By using generating functions, a differential-difference equation can often be reduced to an ordinary differential equation, while the use of Laplace transforms often enables one to reduce a differential-difference equation to a difference equation. A combination of the two tools sometimes facilitates the solution to a differential-difference equation. We illustrate the use of these tools by some examples.

Example 6(a). Solve the differential-difference equation

$$u'_n(t) = u_{n-1}(t), \quad t \geqslant 0, \quad n = 1, 2, \ldots \tag{6.1}$$

given the initial conditions

$$u_0(t) = 1,\ t \geqslant 0 \tag{6.2}$$

and

$$u_0(0) = 1,\ u_n(0) = 0 \text{ for } n \neq 0 \tag{6.2a}$$

Let

$$G(s,t) = \sum_{n=0}^{\infty} u_n(t)s^n \tag{6.3}$$

be the generating function of the sequence $\{u_n(t)\}$; assuming the validity of term by term differentiation, we get

$$\frac{\partial G}{\partial t} = \sum_{n=0}^{\infty} u'_n(t)s^n = u'_0(t) + \sum_{n=1}^{\infty} u'_n(t)s^n$$

$$= \sum_{n=1}^{\infty} u'_n(t)s^n, \text{ since } u_0'(t) = 0 \text{ from (6.2).}$$

Multiplying (6.1) by s^n and adding over for $n = 1, 2, \ldots$, we get

$$\sum_{n=1}^{\infty} u'_n(t)s^n = \sum_{n=1}^{\infty} u_{n-1}(t)s^n$$

or,

$$\frac{\partial G}{\partial t} = sG,$$

which is a partial differential equation. The solution is

$$G(s, t) = A \exp(st),$$

and putting $t = 0$, we get $G(s, 0) = A$.

Again, $G(s, 0) = \sum_{n=0}^{\infty} u_n(0)s^n = u_0(0) + \sum_{n=1}^{\infty} u_n(0)s^n = 1$, from (6.2a).

Hence $A = 1$ and $G(s, t) = \exp(st)$.

Thus,

$$u_n(t) \equiv \text{coefficient of } s^n \text{ in } G(s, t)$$

$$= \text{coefficient of } s^n \text{ in } \exp(st)$$

$$= \frac{t^n}{n!},\ n = 0, 1, 2, \ldots .$$

Example 6(b). Consider the equations

$$p'_n(t) = -\lambda[p_n(t) - p_{n-1}(t)],\ n \geqslant 1,\ \lambda > 0 \tag{6.4}$$

$$p'_0(t) = -\lambda p_0(t) \tag{6.5}$$

with the initial conditions $p_n(0) = 0$ for $n \neq 0$, $p_0(0) = 1$. (6.6)

Let

$$L\{p_n(t)\} = \int_0^\infty \exp(-st)p_n(t)dt \equiv f_n(s) \tag{6.7}$$

be the L.T. of $p_n(t)$; then

$$\begin{aligned} L\{p_n'(t)\} &= sf_n(s) - p_n(0), \text{ by equation (2.6)} \\ &= sf_n(s),\ n \geqslant 1. \end{aligned}$$

Hence taking the L.T. of both sides of (6.4), we have

$$sf_n(s) = -\lambda[f_n(s) - f_{n-1}(s)] \tag{6.8}$$

or, $$f_n(s) = \left(\frac{\lambda}{\lambda+s}\right)f_{n-1}(s),\ n \geqslant 1,$$

which is a difference equation in f_n. Its solution is given by

$$f_{n-1}(s) = c\left(\frac{\lambda}{\lambda+s}\right)^{n-1},\ n \geqslant 1. \tag{6.9}$$

By taking L.T. of (6.5) and using the initial condition $p_0(0) = 1$, we get

$$sf_0(s) - 1 = -\lambda f_0(s)$$

or, $$f_0(s) = 1/(s+\lambda). \tag{6.10}$$

The unknown constant c in (6.9) can be obtained by putting $n = 1$ and using (6.10). We have

$$f_0(s) = c = 1/(s+\lambda).$$

Hence

$$f_{n-1}(s) = \frac{\lambda^{n-1}}{(\lambda+s)^n},\ n \geqslant 1$$

or, $$f_n(s) = \frac{\lambda^n}{(\lambda+s)^{n+1}},\ n \geqslant 0. \tag{6.11}$$

Now, $p_n(t)$ can then be obtained by finding the inverse Laplace transform of $f_n(s)$. From Example 2(b) we find that $f_n(s)$ is the L. T. of

$$\frac{\exp(-\lambda t)(\lambda t)^n}{n!}.$$

Hence

$$p_n(t) = \frac{\exp(-\lambda t)(\lambda t)^n}{n!},\ n = 0, 1, 2, \ldots. \tag{6.12}$$

The solutions of these equations governing Poisson processes will be obtained by an alternative method of generating function in Sec. 4.1 (Theorem 4.1).

Example 6(c). Consider the equations

$$p_n'(t) = -(a+b)p_n(t) + ap_{n-1}(t) + bp_{n+1}(t),\ n \geqslant 1 \tag{6.13}$$

$$p_0'(t) = -ap_0(t) + bp_1(t) \tag{6.14}$$

with initial conditions $p_n(0) = 0,\ n \neq 0$ and $p_0(0) = 1$. (6.15)

Let

$$F_n(s) = \int_0^\infty \exp(-st)p_n(t)dt \quad \text{be the L.T. of } p_n(t).$$

By taking L.T. of both sides of (6.13), we get the difference equation

$$sF_n(s) - p_n(0) = -(a+b)F_n(s) + aF_{n-1}(s) + bF_{n+1}(s),\ n \geqslant 1$$

or,

$$F_n(s) = \frac{a}{a+b+s}F_{n-1}(s) + \frac{b}{a+b+s}F_{n+1}(s). \quad (6.16)$$

Similarly from (6.14), we get

$$sF_0(s) - p_0(0) = -aF_0(s) + bF_1(s)$$

or,

$$F_0(s) = \frac{1}{a+s} + \frac{b}{a+s}F_1(s). \quad (6.17)$$

The L.T. of $p_n(t)$ for $n \geqslant 1$ satisfies a homogeneous difference equation of order two. Its characteristic equation is

$$m = \frac{a}{a+b+s} + \frac{b}{a+b+s}m^2$$

or,

$$bm^2 - (a+b+s)m + a = 0. \quad (6.18)$$

If m_1, m_2 are the two roots of this equation, then the solution of (6.16) is

$$F_n(s) = Am_1{}^n + Bm_2{}^n, \quad n \geqslant 1, \quad (6.19)$$

where A, B are two constants. By putting the value of $F_1(s)$, we get $F_0(s)$ from (6.17).

Thus we get $F_n(s)$, the L.T. of $p_n(t)$ $(n \geqslant 0)$, whence $p_n(t)$ can be determined. This technique due to Conolly (1958) leads to the solution of the $M/M/1$ queueing problem (see Sec 10.3).

1.7 Matrix Analysis

A. Eigenvalues and Eigenvectors

Let $A = (a_{ij})$, where a_{ij} are complex numbers, be an $m \times m$ matrix and let I_m denote the unit matrix of order m. Let $\mathbf{x}$ denote a column vector $\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$ and $\mathbf{x}'$ the row vector $(x_1, \ldots, x_m)$; thus $\mathbf{x} = (x_1, \ldots, x_m)'$.

Definitions. The roots t_i, $i = 1, 2, \ldots, m$, of the equation $\det(A - tI) = 0$ are called the *eigenvalues* (or *spectral values*, *latent* or *characteristic roots*) of the matrix A. The equation $\det\ (A - tI) = 0$ which is of degree m over t is called the *characteristic equation* of A.

The set $T = \{t_i, i = 1, 2, \ldots, m\}$ of eigenvalues is called the *spectrum* of A, and $p(A) = \max\{t_1, \ldots, t_m\}$ is called the *spectral radius* of A.

A non-zero column vector $\mathbf{x}$ which satisfies the equations $(A - t_iI)\mathbf{x} = 0$

is called a *right eigenvector* (latent or characteristic vector) *of A corresponding to* (or associated with) the eigenvalue t_i.

A non-zero row vector $\mathbf{y}'$ which satisfies the equation $\mathbf{y}'(A - t_i I) = 0$ is called a *left eigenvector of A corresponding to* t_i.

Now $(A - t_i I).\mathbf{x} = 0$ has non-zero solution $\mathbf{x}$ iff $\det(A - t_i I) = 0$, that is, iff t_i is an eigenvalue of A. An alternative way of defining eigenvalues and eigenvectors is as follows:

A number t is called an eigenvalue of the matrix A if there exists a non-zero (column) vector $\mathbf{x}$ such that $A\mathbf{x} = t\mathbf{x}$; the non-zero vector $\mathbf{x}$ is called the right eigenvector of A corresponding to t.

The equation $A\mathbf{x} = t\mathbf{x}$ being homogeneous, the eigenvector $\mathbf{x}$ cannot be unique.

It may be noted that the equation $\mathbf{y}'(A - t_i I) = 0$ is equivalent to $(A' - t_i I)\,\mathbf{y} = 0$. This can be immediately shown by using the property that for two matrices X, Y,

$$(XY)' = Y'X'.$$

We state here some simple properties of eigenvalues and eigenvectors (the eigenvalues t_i, $i = 1, \ldots, m$ of A need not necessarily be distinct):

1. Det $A = 0$ iff A has a zero eigenvalue.
2. The matrix A and its transverse A' have the same eigenvalues.
3. Different matrices may have the same eigenvalues. For example,

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \text{ and } \begin{pmatrix} 2 & 0 \\ 3 & 1 \end{pmatrix} \text{ have the same eigenvalues.}$$

4. If a matrix A is of the form

$$A = \begin{pmatrix} A_{11} & 0 \\ B_{11} & A_{22} \end{pmatrix},$$

where A_{11}, A_{22} are square matrices, then t is an eigenvalue of A iff it is an eigenvalue of at least one of the matrices A_{11}, A_{22}.

5. The sum of the diagonal elements (i.e. Σa_{ii}) of A of order m is known as the *trace* of A and is denoted by tr (A). Det (A) and tr (A) are expressible in terms of the eigenvalues t_i as follows:

$$\operatorname{tr}(A) = \sum_{i=1}^{m} t_i, \ \det A = \prod_{i=1}^{m} t_i.$$

6. If $f(A)$ is a polynomial (scalar) in A, then the eigenvalues of $f(A)$ are $f(t_i)$, $i = 1, \ldots, m$. In particular, the eigenvalues of A^n are $(t_i)^n$, $i = 1, \ldots, m$.
7. The eigenvectors form a biorthogonal system. Let t_i, t_j be two distinct eigenvalues with $\mathbf{x}_i$, $\mathbf{x}_j$ and $\mathbf{y}_i'$, $\mathbf{y}_j'$ as the corresponding eigenvectors.

We have

$$A\mathbf{x}_i = t_i\mathbf{x}_i \quad \text{and} \quad \mathbf{y}_j'A = t_j\mathbf{y}_j'$$

and $$t_j(\mathbf{y}_j'\mathbf{x}_i) = (t_j\mathbf{y}_j')\,\mathbf{x}_i = (\mathbf{y}_j'A)\,\mathbf{x}_i$$
$$= \mathbf{y}_j'\,(A\mathbf{x}_i) = \mathbf{y}_j' t_i \mathbf{x}_i = t_i(\mathbf{y}_j'\mathbf{x}_i).$$

Since $t_i \neq t_j$, equality holds iff $\mathbf{y}_j'\mathbf{x}_i = 0$.

B. Nonnegative Matrices

We shall now confine ourselves to square matrices with real elements. A matrix A is nonnegative ($A \geqslant 0$) if all the elements of A are nonnegative, i.e. $A \geqslant 0$ if $a_{ij} \geqslant 0$, and A is positive ($A > 0$) if $a_{ij} > 0$ for all $i, j = 1, 2, \ldots, m$.

Definition. A square matrix in which each row and each column has exactly one entry with value unity, and all others zero is called a *permutation matrix*.

Note that there are $m\,!$ permutation matrices of order m. An $m \times m$ matrix A is said to be *reducible* if there exists a permutation matrix P such that

$$PAP' = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}.$$

where A_{11}, A_{22} are squares matrices of order r and $m - r$ respectively.

A is *irreducible* if no such permutation matrix P exists.

It is to be noted that a transformation of the matrix A of the form $P'AP$ has simply the effect of rearranging the elements of A. The reducibility of a matrix depends on the disposition of the zero and nonzero elements of A. A matrix A with all positive real elements is necessarily irreducible. Further, if the matrix is irreducible, all the off-diagonal entries of any row or column cannot vanish. We state below an elegant result concerning nonnegative matrices due to Perron-Frobenius.

Theorem 1.12. Let A be a nonnegative and irreducible matrix. Then

(1) A has a positive eigenvalue, r, equal to its spectral radius $p(A)$,

(2) r is a simple eigenvalue of A, i.e. r is a simple root of the characteristic equation,

(3) there is a positive right eigenvector associated with r, and

(4) r increases when any entry of A increases.

Perron obtained the result in 1907 for positive matrices; it was later extended by Frobenius in 1912 to nonnegative matrices. If A is positive, (when it is necessarily irreducible) Perron showed that there is only one eigenvalue of modulus r. If A is nonnegative and irreducible, then the number of eigenvalues having modulus equal to $r\,(= p(A))$ is h, where $h \geqslant 1$.

Definition. An nonnegative irreducible matrix is called *primitive*, if $h = 1$, and *imprimitive* (or *cyclic*) of index h, if $h > 1$.

A positive matrix is necessarily primitive.

A nonnegative irreducible matrix A may be primitive or imprimitive. Suppose that A is cyclic of index h and t_i, $i = 1, 2, \ldots, h$ are the h eigenvalues of modulus r, i.e. $t_i = r = p(A)$, then $t_i = w_i r$, $i = 1, \ldots, h$, where w_i's are the h distinct roots of unity.

C. Spectral Representation of a Matrix

Consider an $m \times m$ matrix A with m distinct *non-zero* eigenvalues t_i and $\mathbf{x}_i$, $\mathbf{y}_i'$ as the right and left eigenvectors, $i = 1, 2, \ldots, m$, of A. Let $D = (d_{ij}), d_{ij} = 0, i \neq j, d_{ii} = t_i$ be a diagonal matrix and let $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $Y = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ be the matrices formed with the right and left eigenvectors respectively of A. From $\mathbf{y}_j'A = t_j\mathbf{y}_j'$ and $A\mathbf{x}_j = t_j\mathbf{x}_j$, we get $Y'A = DY'$ and $AX = XD$ respectively. Now suppose X^{-1} exists, then $AX = XD$ implies $X^{-1}(AX)X^{-1} = X^{-1}(XD)X^{-1}$, or $X^{-1}A = DX^{-1}$. Thus the rows of X^{-1} are the left eigenvectors of A.

Particular case: Suppose that $X^{-1} = Y'$; this implies that $XY' = I$, or $Y'X = I$. We may thus choose suitable multiplicative factors to get $\mathbf{y}'\mathbf{x} = 1$. Since $\mathbf{y}'\mathbf{x} = 0$, $i \neq j$, we may write $\mathbf{y}_j'\mathbf{x}_i = \delta_{ij}$. From $AX = XD$ we get $A = XDX^{-1} = XDY'$. Further $A^2 = (XDX^{-1})(XDX^{-1}) = XD^2X^{-1} = XD^2Y'$ and $A^n = XD^nY'$ (for any positive integer $n \geqslant 1$), where D^n is a diagonal matrix with elements t_i^n, $i = 1, 2, \ldots, m$, along the main diagonal, the off-diagonal elements being all zero. On expanding, we get

$$A^n = \sum_{j=1}^{m} t_j^n \mathbf{x}_j \mathbf{y}_j'.$$

General case: Suppose that the eigenvectors $\mathbf{x}$ and $\mathbf{y}'$ are arbitrarily chosen; then we can write

$$\mathbf{y}_j'\mathbf{x}_i = \delta_{ij}/c_j, \quad \text{or} \quad c_j = 1/(\mathbf{y}_j'\mathbf{x}_j)$$

$$A^n = \sum_{j=1}^{m} t_j^n c_j \mathbf{x}_j \mathbf{y}_j', \; n \geqslant 1.$$

The products $c_j\mathbf{x}_j\mathbf{y}_j' = B_j$ are matrices (called *constituent matrices*) and the above gives an expression of A^n as a sum of m such matrices.

We immediately find the following properties of constituent matrices:

(i) $B_iB_j = 0$, for $i \neq j$;

(ii) $B_j^2 = B_j$ (B_j's are idempotent);

and (iii) $\sum_{j=1}^{m} B_j = I_m$.

The representation $A = XDY'$ is known as the *spectral* (*or canonical*) representation of A.

Example 7(a). Consider the matrix

$$A = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}$$

The solutions of the equation $\det(A - tI) = 0$ are 1 and 5, i.e. $t_1 = 1$, $t_2 = 5$ are the eigenvalues. The right eigenvector $\mathbf{x}_1$ (where $\mathbf{x}_1' = (x_{11}, x_{12})$) corresponding to $t_1 = 1$ is given as the solution of $A\mathbf{x}_1 = \mathbf{x}_1$, i.e.

$$2x_{11} + x_{12} = x_{11} \quad \text{and} \quad 3x_{11} + 4x_{12} = x_{12}.$$

Solving we get,

$$\mathbf{x}_1' = (x_{11}, x_{12}) = (1, -1).$$

Corresponding to $t_2 = 5$, to right eigenvector $\mathbf{x}_2$ (where $\mathbf{x}_2' = (x_{21}, x_{22})$) is given as the solutions of $A\mathbf{x}_2 = 5\mathbf{x}_2$, i.e.

$$2x_{21} + x_{22} = 5x_{21} \quad \text{and} \quad 3x_{21} + 4x_{22} = 5x_{22}.$$

We get,

$$\mathbf{x}_2' = (x_{21}, x_{22}) = (1, 3).$$

The left eigenvectors $\mathbf{y}_i' = (y_{i1}, y_{i2})$ can be similarly obtained as solutions of $\mathbf{y}_i' A = t_i \mathbf{y}_i'$, $i = 1, 2$. We get

$$\mathbf{y}_1' = (y_{11}, y_{12}) = (-3, 1)$$

and

$$\mathbf{y}_2' = (y_{21}, y_{22}) = (1, 1).$$

Thus,

$$c_1 = 1/(\mathbf{y}_1'\mathbf{x}_1) = -\tfrac{1}{4}, \; c_2 = 1/(\mathbf{y}_2'\mathbf{x}_2) = \tfrac{1}{4},$$

$$\mathbf{x}_1\mathbf{y}_1' = \begin{pmatrix} -3 & 1 \\ 3 & -1 \end{pmatrix}, \; \mathbf{x}_2\mathbf{y}_2' = \begin{pmatrix} 1 & 1 \\ 3 & 3 \end{pmatrix};$$

$$A = \sum_{i=1}^{2} t_i B_i, \; A^r = \sum_{i=1}^{2} t_i^r B_i^r \; (r \geqslant 1),$$

where $B_i = c_i \mathbf{x}_i \mathbf{y}_i'$.

Verify that (i) $\mathbf{x}$, $\mathbf{y}$ are biorthogonal,

$$\text{(ii)} \quad B_i^2 = B_i, \; B_i B_j = 0, \; i \neq j, \; \sum_{i=1}^{2} B_i = I_2.$$

D. Stochastic Matrix

Definition. A square matrix with nonnegative elements and unit row sums is called a *stochastic* (*Markov*) matrix. It is *doubly stochastic*, if, in addition, it has unit column sums.

$P = (p_{ij})$ is stochastic iff $p_{ij} \geqslant 0$, $\sum_j p_{ij} = 1$ for all i, j;

P is doubly stochastic iff $p_{ij} \geqslant 0$, $\sum_j p_{ij} = 1$, $\sum_i p_{ij} = 1$ for all i, j.

If P is a stochastic matrix, so also is P^n, n being a positive integer. Stochastic matrices (which are nonnegative) have some very interesting properties which are useful in the study of Markov chains. Some of these are given below.

Theorem 1.13. The matrix D is stochastic iff it has an eigenvalue 1 and the corresponding right eigenvector $\mathbf{x}$ is a column vector with all the elements equal to unity, i.e. $\mathbf{x}' = (1, 1, \ldots, 1)$. Further $p(P) = 1$, i.e. the spectral radius of a stochastic matrix is 1.

If the matrix P is irreducible, then there can be one and only one eigenvalue equal to 1 (Theorem 1.12). However, there can be $h (> 1)$ eigenvalues of modulus 1 if P is cyclic of order h. If it is primitive ($h = 1$), then there can be no other eigenvalue of modulus 1. From this we get that $\lim\limits_{n\to\infty} P^n \neq 0$.

Theorem 1.14. Suppose that P is an irreducible stochastic matrix. Then $\lim\limits_{n\to\infty} P^n$ exists iff P is primitive.

Remark: The hypothesis that P is irreducible is not necessary; the limit P^n exists when P is a primitive stochastic matrix.

These properties of stochastic matrices will be used in the study of Markov chains (Sec. 3.4).

EXERCISES

1.1. (a) Find the generating function of the *Fibonacci numbers* $\{f_n\}$ (defined in Example 4(a)).

(b) Let X be a random variable denoting the number of tosses required to get two consecutive heads when a fair coin is tossed. Show that the p.g.f. of X is $(s^2/4)\{1 - s/2 - (s/2)^2\}^{-1}$.

1.2. Let Y have the distribution of the geometric form (modified geometric or decapitated geometric) given by

$$\Pr\{Y = k\} = q^{k-1}p,\ k = 1, 2, 3, \ldots.$$

Show that the p.g.f. of Y is $ps/(1 - sq)$, and that

$$E(Y) = 1/p \text{ and } \operatorname{var}(Y) = q/p^2.$$

(The number of trials Y upto and including the first success in a sequence of Bernoulli trials with probability p of success has the above distribution; again $Y = X + 1$, where X has geometric distribution considered in Example 1(*c*).)

1.3. Let X be a random variable such that

$$\Pr\{X = k\} = p_k,\ \Pr\{X > k\} = q_k = \sum_{r=k+1}^{\infty} p_r,\ k \geq 0.$$

If $\quad P(s) = \sum_{k=0}^{\infty} p_k s^k \quad$ and $\quad Q(s) = \sum_{k=0}^{\infty} q_k s^k,$

show that $(1-s)\,Q(s) = 1 - P(s)$ and that $E(X) = Q(1)$.

1.4. Find the mean and variance of the negative binomial (NB) distribution considered in Example 1(g). Obtain the p.g.f. of the sum of m i.i.d. NB variables and find its mean and variance.

1.5. (a) Let X have a zero-truncated (or decapitated) Poisson distribution with zero class missing, i.e.

$$p_k = \Pr\{X = k\} = \frac{(e^a - 1)^{-1}\, a^k}{k!}, \quad k = 1, 2, 3, \ldots.$$

Show that the p.g.f. of X is given by $P(s) = (e^a - 1)^{-1}(e^{as} - 1)$.

Verify that $P(1) = \sum\limits_{k=1}^{\infty} p_k = 1$; show that $E(X) = ae^a/(e^a - 1)$.

(b) If X, Y are two independent zero-truncated Poisson variates, find the p.g.f. of the sum $(X + Y)$ and the distribution $\Pr\{X + Y = n\}$, $n = 2, 3, \ldots$ when (i) X, Y are identical with the same parameter a, (ii) X, Y have parameters a_1, a_2 (see Medhi, *Metrika*, **20**, 215–218 (1973)).

(c) Find the p.g.f. of the sum $S_n = X_1 + \ldots + X_n$ of n independent and identical zero-truncated Poisson distributions. Find $E(S_n)$ and $\Pr\{S_n = m\}$, $m = n, n+1, n+2, \ldots$.

1.6. Let $S_N = X_1 + \ldots + X_N$, where X_i are i.i.d. random variables and N is a random variable independent of X_i's. Show that

$$\operatorname{var}(S_N) = E\{N\}\operatorname{var}(X_i) + \operatorname{var}(N)\{E(X_i)\}^2.$$

and that

$$\frac{\operatorname{var}(S_N)}{E(S_N)} = \frac{\operatorname{var}(X_i)}{E(X_i)} + E(X_i)\,\frac{\operatorname{var}(N)}{E(N)}.$$

Show that for a compound Poisson or compound geometric distribution,

$$\frac{\operatorname{var}(S_N)}{\{E(S_N)\}^2} \geqslant k\,\frac{\operatorname{var}(X_i)}{\{E(X_i)\}^2},$$

where k is simply related to the compounding parameter.

1.7. Let $S_N = X_1 + X_2 + \ldots + X_N$, where N has Poisson distribution with mean a.

(a) If X_i have i.i.d. geometric distribution with mean q/p, find (i) the p.g.f. of S_N, and (ii) $E(S_N)$ and $\operatorname{var}(S_N)$.

(b) If X_i have i.i.d. Bernoulli distributions with $\Pr\{X_i = 1\} = p$, and $\Pr\{X_i = 0\} = 1 - p = q$, show that S_N has Poisson distribution with mean ap. Show that the joint distribution of S_N and N has probability mass function

$$\Pr\{N = n, S_N = y\} = \frac{e^{-a}a^n p^y q^{n-y}}{y!(n-y)!}$$

for $\quad n = 0, 1, 2, \ldots, y = 0, 1, 2, \ldots, n$

and that the conditional distribution of N given $S_N = y$ has probability mass function

$$\Pr\{N=n \mid S_N=y\}=\frac{e^{-aq}(aq)^{n-y}}{(n-y)!},\ n\geqslant y.$$

Show that $\operatorname{cov}(N, S_N)=ap.$

(c) If X_i have i.i.d. Poisson distributions with mean b, show that the joint distribution of S_N, and N has mass function

$$\Pr\{N=n, S_N=y\}=\frac{\exp\{-(a+bn)\}\, a^n (nb)^y}{n!\, y!}$$

for $n=0, 1, 2, \ldots, y=0, 1, \ldots.$

Show that $E(S_N)=ab$ and $\operatorname{var}(S_N)=ab(1+b)$.

(Accident models have been constructed and fitted to actual data involving traffic accidents (Leiter and Hamdan, *Int. Stat. Rev.* **41**(1973) 87–100) taking N as the number of accidents; X_i (in (b)) as the indicator variable indicating whether the ith accident results in a fatality or not; X_i (in (c)) as the number of persons fatally injured in the ith accident.)

1.8. Let $Z=X_1+\ldots+X_N$, where $\{N\}$ has zero-truncated Poisson distribution (defined in 5(a) above) (N is independent of X_i) and X_i have i.i.d. exponential distribution with mean $1/\lambda$. Show that the L.T. $f^*(s)$ of Z is

$$f^*(s)=\{\exp(a\lambda/(\lambda+s))-1\}/(e^a-1),$$

and that

$$E(Z)=ae^a/\{\lambda(e^a-1)\}.$$

It may be further verified that

$$f(t)=\frac{\exp(-\lambda t)}{\exp(a)-1}\sqrt{\left(\frac{a\lambda}{t}\right)}\, I_1(2\sqrt{(\lambda a t)})$$

is the probability density function of S_N. (Show that L.T. of $f(t)$ is $f^*(s)$.)

($I_r(x)$ is the modified Bessel function of order $r(\geqslant -1)$ given by

$$I_r(x)=\sum_{k=0}^{\infty}\frac{1}{k!\,\Gamma(k+r+1)}(x/2)^{2k+r}.$$

(The above result can be used in Queueing theory in a model of phase-type service, where the r.v. N denotes the number of phases associated with a customer's service, and the i.i.d. r.v. X_i denotes the duration of service time at the ith phase.)

The Bessel function density $f(t)$ is quite similar, for the value of the parameter a of 10 or over, to the gamma density $f_{\lambda,a}(t)$. However $f(t)$ gives better fit to a large number of empirical distributions (see Gaver *Opns. Res.* **2** (1954) 147).

1.9. Suppose that X_i, $i=1, 2$ are two independent geometric variables with $\Pr\{X_i=k\}=pq_i^k$, $i=1, 2$, $k=0, 1, 2, \ldots$. Find the bivariate generating function $P(s_1, s_2)$ of the pair (X_1, X_2) and from the form of the generating function for the sum $S=X_1+X_2$, verify that

$$\Pr\{S=k\} = \sum_{r=0}^{k} q_1{}^r p_1 q_2{}^{k-r} p_2, \quad k = 0, 1, 2 \,. \,. \,.$$

1.10. Find the Laplace transforms of:

(i) $\sin at$; (ii) $a \exp(-2t)$; (iii) $(t^2+1)^2$; (iv) $t^3 \exp(-3t)$; (v) $e^{-t}\cos 2t$; (vi) $t^n e^{at} (n > -1)$;

(vii) $\dfrac{\{\exp(-at)\}(at)^n}{n!}$ $(n = 0, 1, 2, \ldots)$; (viii) $\dfrac{a \exp\{-(a^2/4t)\}}{2\sqrt{(\pi t^3)}}$;

(ix) $\sqrt{(t^r)} I_r(2\sqrt{t})$; (x) $\{\exp(-t-1)\}\sqrt{(t^r)} I_r(2\sqrt{t})$;

(xi) $\dfrac{\exp(-\lambda t)}{e^a - 1}\sqrt{\left(\dfrac{\lambda a}{t}\right)} I_1(2\sqrt{(\lambda a t)})$.

1.11. If $L\{f(t)\} = F(s)$,

show that

(i) $L\{tf(t)\} = -F'(s)$,

(ii) $L\{t^2 f(t)\} = (-1)^2 F''(s)$,

(iii) $L\{f''(t)\} = s^2 F(s) - sf(0) - f'(0)$.

($f'(t)$ denotes differentiation with respect to t and $F'(s)$ denotes differentiation with respect to s).

(The generalisations of (i) and (ii) for higher powers of t and of (iii) for higher derivatives can also be obtained.)

1.12. The non-negative valued random variable X is said to have a *modified exponential distribution* with an non-zero probability mass at '0', if its distribution function is

$$F(x) = \Pr\{X \leqslant x\} = 1 - k \exp(-\lambda x),$$
$$x \geqslant 0, \quad \lambda > 0, \quad 0 < k < 1.$$

Find (i) the frequency density of X, (ii) $E(X)$, and (iii) L.T. of X.

1.13. Show that the sum of two independent gamma variables having parameters λ, k and λ, l is a gamma variable with parameters $\lambda, k+l$.

1.14. Find the Laplace transform of the variable X (having triangular distribution) with density function

$$\begin{aligned} f(x) &= x, & 0 \leqslant x \leqslant 1 \\ &= 2 - x, & 1 \leqslant x \leqslant 2. \end{aligned}$$

Hence show that the triangular distribution can be obtained as the sum of two independent random variables each with a rectangular distribution in (0, 1).

1.15. Solve the difference equations

(i) $f(n+2) - f(n) = 0$,

(ii) $U_{n+2} + 2U_{n+1} + U_n = 0$,

(iii) $f(n+3) - 4f(n+2) + 5f(n+1) - 2f(n) = 0.$

1.16. Solve the equation

$$U_n = pU_{n+1} + qU_{n-1},\ 1 \leqslant n \leqslant a-1,\ 0 < p = 1-q < 1$$

given that

$$U_0 = 1,\ U_a = 0,\ (p \neq q).$$

By taking the limit of U_n as $p/q \to 1$, find the solution of

$U_n = \frac{1}{2}(U_{n+1} + U_{n-1})$ (see Example 5(a)).

1.17. If $p_n(t)$, $n = 0, 1, 2 \ldots$ represents a probability distribution and $f_n(s)$ is the L.T. of $p_n(t)$, then show that

$$\sum_{n=0}^{\infty} f_n(s) = 1/s.$$

1.18. Apply the method of Laplace transform (employed in Example 6(b)) to solve the equation of Example 6(a), viz. (6.1) with the given initial conditions (6.2) and (6.2a).

1.19. Consider the differential-difference equations (6.13) and (6.14) with initial conditions (6.15) (see Example 6(c)).
If

$$G(z,t) = \sum_{n=0}^{\infty} p_n(t) z^n$$

show that

$$z\frac{\partial G(z,t)}{\partial t} = (1-z)\{(b-az)G(z,t) - bp_0(t)\}.$$

If $p_n{}^*(s) = L\{p_n(t)\}$ and $G^*(z,s) = L\{G(z,t)\}$,

then show that

$$G^*(z,s) = \frac{z - b(1-z)\,p_0{}^*(s)}{sz - (1-z)(b-az)}.$$

(These relations occur in the study of transient solution of $M/M/1$ queue (Bailey 1954)); see Sec 10 3.)

1.20. Suppose that $P_n(t)$ satisfies the differential-difference equations

$$P_n'(t) = -(a+b)P_n(t) + aP_{n-1}(t) + bP_{n+1}(t),\ n \geqslant 2$$
$$P_1'(t) = -(a+b)P_1(t) + bP_2(t)$$
$$P_0'(t) = bP_1(t)$$

with initial conditions $P_1(0) = 1$, $P_n(0) = 0$ for $n \neq 1$.

Show that the generating function $G(z,t) = \sum_{n=0}^{\infty} P_n(t)z^n$ satisfies

$$z\frac{\partial G(z,t)}{\partial t} = (1-z)(b-az)\{G(z,t) - P_0(t)\};$$

show that L.T. $G^*(z,s)$ of $G(z,t)$ is given by

$$G^*(z,s) = \frac{z^2 - (1-z)(b-az)P_0^*(s)}{sz - (1-z)(b-az)}.$$

(These relations lead to busy period distribution of $M/M/1$ queue Bailey (1954).)

1.21. Find the eigenvalues and the left eigenvectors of the matrix

$$P = \begin{pmatrix} .5 & .5 & 0 \\ .5 & 0 & .5 \\ 0 & 1 & 0 \end{pmatrix}.$$

1.22. Examine whether the following matrices are primitive :

$$\text{(i)} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{(ii)} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$\text{(iii)} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix} \quad \text{(iv)} \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix}$$

$$\text{(v)} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

1.23. If the two $m \times m$ matrices P and Q are stochastic, then show that the matrix

(i) PQ is stochastic

(ii) $aP + (1-a)Q$ is stochastic, for $0 \leqslant a \leqslant 1$.

Is the product of two doubly stochastic matrices doubly stochastic?

1.24. Probability vector : a vector $\mathbf{x}' = (x_1, \ldots, x_i, \ldots, x_m)$, such that

$$x_i \geqslant 0, \sum_{i=1}^{m} x_i = 1 \text{ is called a probability vector.}$$

Show that an $m \times m$ matrix M is stochastic iff $x'M$ is a probability vector whenever x' is a probability vector. (This is a characterization of stochastic matrices.)

1.25. Prove that a doubly stochastic $m \times m$ matrix A may be written as

$$A = \Sigma c_r P_r,\ r = 1, \ldots, m!,\ c_r \geqslant 0,\ \Sigma c_r = 1,$$

where P_r are permutation matrices of order m.

REFERENCES

N.T.J. Bailey, "A continuous time treatment of a simple queue using generating functions," *J.R.S.S.*, B, **16** 288-291 (1954).

R. Bellman and K.L. Cooke, *Differential-Difference Equations*, Academic Press, New York, (1963).

B.W. Conolly, "A difference equation technique applied to the simple queue", *J.R.S.S.*, B, **20**, 165-167 (1958).

W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. I, 3 ed. (1968), Vol. II, (1965), Wiley, New York.

C. Jordan, *Calculus of Finite Differences*, Chelsea, New York, (1950).

P. Lancaster, *Theory of Matrices*, Academic Press, New York, (1969).

E. Parzen, *Modern Probability Theory and its Applications*, New York, (1960).

M.G. Smith, *Laplace Transform Theory*, Van Nostrand, London, (1966).

D.V. Widder, *The Laplace Transform*, Princeton University Press, Princeton, N. J. (1941).

CHAPTER 2

STOCHASTIC PROCESSES

2.1 Notion of Stochastic Processes

Since the last century there have been marked changes in the approach to scientific enquiries. There has been greater realisation that probability (or non-deterministic) models are more realistic than deterministic models in many situations. Observations taken at different time points rather than those taken at a fixed period of time began to engage the attention of probabilists. This led to a new concept of indeterminism—indeterminism in dynamic studies. This has been called "dynamic indeterminism" (Neyman (1960), traces the transitions). The period of dynamic indeterminism began roughly with the work of Mendel (1822-1884). The physicists played a leading role in the development of dynamic indeterminism. Many a phenomenon occurring in physical and life sciences are studied now not only as a random phenomenon but also as one changing with time or space. Similar considerations are also made in other areas, such as, social sciences, engineering and management and so on. The scope of applications of random variables which are functions of time or space or both has been on the increase.

Families of random variables which are functions of say, time, are known as stochastic processes (or random processes, or random functions). A few simple examples are given as illustrations.

Example 1(a). Consider a simple experiment like throwing a true die. (i) Suppose that X_n is the outcome of the nth throw, $n \geqslant 1$. Then $\{X_n, n \geqslant 1\}$ is a family of random variables such that for a distinct value of $n\ (= 1, 2, \ldots)$, one gets a distinct random variable X_n. $\{X_n, n \geqslant 1\}$ constitutes a stochastic process, known as Bernoulli process. (ii) Suppose that X_n is the number of sixes in the first n throws. For a distinct value of $n = 1, 2, \ldots$, we get a distinct binomial variable X_n; $\{X_n, n \geqslant 1\}$ which gives a family of random variables is a stochastic process. (iii) Suppose that X_n is the maximum number shown in the first n throws. Here $\{X_n, n \geqslant 1\}$ constitutes a stochastic process.

Example 1(b). Consider that there are r cells and an infinitely large number of identical balls and that balls are thrown at random, one by one, into the cells, the ball thrown being equally likely to go into any one of the cells. Suppose that X_n is the number of occupied cells after n throws. Then $\{X_n, n \geqslant 1\}$ constitutes a stochastic process.

Example 1(c). Consider a random event occurring in time, such as, number of telephone calls received at a switchboard. Suppose that $X(t)$ is the random variable which represents the number of incoming calls in an interval $(0, t)$ of duration t units. The number of calls within a fixed interval of *specified duration*, say, one unit of time, is a random variable $X(1)$ and the family $\{X(t), t \in T\}$ constitutes a stochastic process $(T = [0, \infty))$.

2.2 Specification of Stochastic Processes

The set of possible values of an individual random variable X_n $(X(t))$ of a stochastic process $\{X_n, n \geqslant 1\}$ $(X(t), t \in T)$ is known as its state space. The *state space* is discrete if it contains a finite or a denumerable infinity of points; otherwise, it is continuous. For example, if X_n is the total number of sixes appearing in the first n throws of a die, the set of possible values of X_n is the finite set of nonnegative integers $0, 1, \ldots, n$. Here the state space of X_n is discrete. We can write $X_n = Y_1 + \ldots + Y_n$, where Y_i is a discrete r. v. denoting the outcome of the ith throw and $Y_i = 1$ or 0 according as the ith throw shows six or not. Secondly, consider $X_n = Z_1 + \ldots + Z_n$, where Z_i is a continuous r. v. assuming values in $[0, \infty)$. Here the set of possible values of X_n is the interval $[0, \infty)$, and so the state space of X_n is continuous.

In the above two examples we assume that the parameter n of X_n is restricted to the nonnegative integers $n = 0, 1, 2, \ldots$. We consider the state of the system at distinct time points $n = 0, 1, 2, \ldots$, only. Here the word *time* is used in a wide sense. We note that in the first case considered above the "time n" implies throw number n.

On the other hand, one can visualise a family of random variables $\{X_t, t \in T\}$ (or $\{X(t), t \in T\}$) such that the state of the system is characterized at every instant over a finite or infinite interval. The system is then defined for a continuous range of times and we say that we have a family of r.v. in *continuous* time. A stochastic process in continuous time may have either a discrete or a continuous state space. For example, suppose that $X(t)$ gives the number of incoming calls at a switchboard in an interval $(0, t)$. Here the state space of $X(t)$ is discrete though $X(t)$ is defined for a continuous range of time. We have a process in continuous time having a discrete state space. Suppose $X(t)$ represents the maximum temperature at a particular place in $(0, t)$, then the set of possible values of $X(t)$ is continuous. Here we have a system in continuous time having a continuous state space.

So far we have assumed that the values assumed by the r.v. X_n (or $X(t)$) are one-dimensional, but the process $\{X_n\}$ (or $\{X(t)\}$) may be multi-dimensional. Consider $X(t) = (X_1(t), X_2(t))$, where X_1 represents the maximum and X_2 the minimum temperature at a place in an interval of time $(0, t)$. We have here a two-dimensional stochastic process in continuous time having continuous state space. One can similarly have multi-dimensional processes. A one-dimensional process can be classified into one of the following four types of processes:

(i) Discrete time, discrete state space
(ii) Discrete time, continuous state space
(iii) Continuous time, discrete state space
(iv) Continuous time, continuous state space.

All the four types may be represented by $\{X(t),\ t \in T\}$. In case of discrete time, the parameter generally used is n, i.e. the family is represented by $\{X_n,\ n = 0, 1, 2, \ldots\}$. In case of continuous time both the symbols $\{X_t,\ t \in T\}$ and $\{X(t),\ t \in T\}$ (where T is a finite or infinite interval) are used. The parameter t is usually interpreted as time, though it may represent such characters as distance, length, thickness and so on. Some authors call the discrete parameter family a stochastic *sequence*, and the continuous parameter family a stochastic *process*.

We shall use here the notation $\{X(t),\ t \in T\}$ both in the cases of discrete and continuous parameters and shall specify T, whenever necessary.

Relationship

In some of the cases, the r.v. X_n, i.e. members of the family $\{X_n\}$ are mutually independent (as in Example 1 (a)), but more often they are not. We generally come across processes whose members are mutually dependent. The relationship among them is often of great importance.

The nature of dependence could be infinitely varied. Here dependence of some special types, which occurs quite often and is of great importance, will be considered. We may broadly describe some stochastic processes according to the nature of dependence relationship existing among the members of the family.

Processes with independent increments

If for all $t_1, \ldots, t_n$, $t_1 < t_2 < \ldots < t_n$, the random variables

$$X(t_2)-X(t_1),\ X(t_3)-X(t_2),\ \ldots,\ X(t_n)-X(t_{n-1})$$

are independent, then $\{X(t),\ t \in T\}$ is said to be a process with independent increments.

Suppose that we wish to consider the discrete parameter case. Consider a process in discrete time with independent increments. Writing

$$T = \{0, 1, 2, \ldots\},\ t_i = i-1,\ X(t_i) = X_{i-1},$$

$$Z_i = X_i - X_{i-1},\ i = 1, 2, \ldots \quad \text{and } Z_0 = X_0,$$

we have a sequence of independent random variables $\{Z_n,\ n \geqslant 0\}$.

Markov process. If $\{X(t),\ t \in T\}$ is a stochastic process such that, given the value $X(s)$, the values of $X(t),\ t > s$, do not depend on the values of $X(u),\ u < s$, then the process is said to be a Markov process.

A definition of such a process is given below.

If, for, $t_1 < t_2 < \ldots < t_n < t$

$$\begin{aligned}\Pr\{a \leqslant X(t) \leqslant b \mid X(t_1) = x_1, \ldots, X(t_n) = x_n\} \\ = \Pr\{a \leqslant X(t) \leqslant b \mid X(t_n) = x_n\}\end{aligned}$$

the process $\{X(t), t \in T\}$ is a Markov process.

A discrete parameter Markov process is known as a Markov chain.

2.3 Stationary Processes

If for arbitrary $t_1, \ldots, t_n$, the joint distribution of the vector random variables

$$(X(t_1), X(t_2), \ldots, X(t_n)), \quad \text{and} \quad (X(t_1 + h), \ldots, X(t_n + h))$$

are the same for all $h > 0$, then the stochastic process $\{X(t), t \in T\}$ is said to be *stationary* of order n. It is *strictly stationary* if it is stationary of order n for any integer n.

Stationarity of a process implies that the probabilistic structure of the process is invariant under translation of the time axis. Many processes encountered in practice exhibit such a characteristic. Now for such a strictly stationary process, the statistical properties of the associated joint distribution of $\{X(t_1 + h), \ldots, X(t_n + h)\}$ will obviously be independent of h. This will hold, in particular, in case of the moments of the distribution, when these exist.

If the mean of the process $\{X(t)\}$ exists, then $E\{X(t)\}$ must be equal to $E\{X(t + h)\}$ for any h, so that $E\{X(t)\}$ must be a constant m, independent of t. Without loss of generality, we shall assume that m is zero. If, moreover, the covariance function $C(s, t)$ (or $C_{s,t}$) exists, then

$$\begin{aligned} C_{s,t} &= \operatorname{cov}\{X(t), X(s)\} = E\{X(t)\, X(s)\} \\ &= E\{X(t + h)\, X(s + h)\}, \text{ for any } h \\ &= E\{X(t-s)\, X(0)\}. \end{aligned}$$

This shows that $C_{s,t}$ is a function of the time difference $|t-s|$.

Higher moments, when they exist, can also be considered. Many of the important questions relating to a stochastic process can be adequately answered on the basis of the knowledge of the first two moments of the process. Accordingly, another notion of stationarity, involving existence and stationarity of the moments of order one and two, is also considered.

A process $X(t)$ with finite second order moments is called *covariance stationary* or *weakly stationary* or *wide sense stationary* (or simply *stationary*), if its mean $E\{X(t)\} = m$ is independent of t, and its covariance function

$$C(s,t) = E\{X(t)\, X(s)\}$$

depends only on the difference $|t-s|$, for all t, s. We shall also use the notation $C_{s,t}$ for $E\{X(t)\ X(s)\}$, and $m(t)$ for $E\{X(t)\}$.

Note that a strictly stationary process will not necessarily be a weakly stationary process, nor will a weakly stationary process be necessarily strictly stationary. A process which is not stationary (in any sense) is said to be *evolutionary*. The processes considered here will be assumed to be real throughout. We now define a class of stochastic processes in terms of an important specific distribution—normal distribution.

Gaussian processes. If the distribution of $(X(t_1), \ldots, X(t_n))$ for all $t_1, \ldots, t_n$ is multivariate normal, then $X(t)$ is said to be a *Gaussian* process.

Such processes derive their importance from the fact that there are many physical processes which are approximately normal, e.g. the process describing atmospheric turbulence. Further, such processes have some very useful and simple properties. We state below such a property.

If a Gaussian process $X(t)$ is stationary, then it is strictly stationary.

The explanation is simple. The multivariate normal distribution of $(X(t_1), \ldots, X(t_n))$ is completely determined by its mean vector $(\mu_1, \ldots, \mu_n)$ (where $\mu_i = E\{X(t_i)\}$ and the variance-covariance matrix $(C(i, j))$, whose elements are $C(i, j) = \text{cov}\,\{X(t_i),\ X(t_j)\},\ i, j = 1, 2, \ldots, n$.
If the Gaussian process is covariance stationary, then $E\{X^2(t_i)\}$, $E\{X^2(t_i)\}$ are finite and the covariance function $C(i, j)$ is a function only of the difference $i-j$.

Example 3(a). Let $X_n,\ n \geqslant 1$ be uncorrelated r.v.'s with mean 0 and variance 1. Then

$$\begin{aligned} C(n, m) = \text{cov}\{X_n,\ X_m\} = E\{X_n X_m\} &= 0, \qquad m \neq n \\ &= 1, \qquad m = n \end{aligned}$$

and so $\{X_n,\ n \geqslant 1\}$ is covariance stationary.

If X_n are also identically distributed, then $\{X_n$ is strictly stationary.

Example 3(b). Poisson process. Consider the process $\{X(t), t \in T\}$ with

$$\Pr\{X(t) = n\} = \exp\,(-at)\,(at)^n/n!,\ a > 0,\ n = 0, 1, 2, \ldots.$$

We see that

$$m(t) = at, \quad \text{var}\,\{X(t)\} = at$$

are functions of t. The process is evolutionary. The distribution of the process is functionally dependent on t.

Example 3(c). Consider the process

$$X(t) = A_1 + A_2 t,$$

where A_1, A_2 are independent r.v.'s with $E(A_i) = a_i$, var $(A_i) = \sigma_i^2$, $i = 1, 2$. We have

$$m(t) = a_1 + a_2 t.$$

$$\begin{aligned} E\{X(t)\, X(s)\} &= E\{(A_1 + A_2 t)\,(A_1 + A_2 s)\} \\ &= E(A_1^2) + (s + t)\, E(A_1 A_2) + ts\, E(A_2^2) \\ &= \sigma_1^2 + a_1^2 + (s + t) a_1 a_2 + ts\,(\sigma_2^2 + a_2^2). \end{aligned}$$

Thus

$$E\{X(t)\}^2 = \sigma_1^2 + a_1^2 + 2t\, a_1 a_2 + t^2\,(\sigma_2^2 + a_2^2),$$

$$\text{var}\,\{X(t)\} = \sigma_1^2 + t^2 \sigma_2^2,$$

and

$$C(s, t) = \text{cov}\,\{X(t), X(s)\} = \sigma_1^2 + ts\sigma_2^2.$$

The process is evolutionary.

Example 3(d). Consider the process

$$X(t) = A \cos \omega t + B \sin \omega t,$$

where A, B are uncorrelated r. v.'s each with mean 0 and variance 1 and ω is a positive constant.

We have

$$m(t) = 0,$$

$$\begin{aligned} E\{X(t)X(s)\} &= E\{(A \cos \omega t + B \sin \omega t)\,(A \cos \omega s + B \sin \omega s)\} \\ &= \cos \omega t \cos \omega s + \sin \omega t \sin \omega s \\ &= \cos\,(s - t)\omega. \end{aligned}$$

Thus

$$E\{(X(t))^2\} = \text{var}\,\{X(t)\} = 1,$$

$$C(s, t) = \text{cov}\,\{X(t), X(s)\} = \cos\,(s - t)\omega.$$

Here the first two moments are finite and the covariance function is a function of $(s - t)$. Thus the process is covariance stationary.

Example 3(e). Consider the process $\{X(t),\ t \in T\}$ whose probability distribution, under a certain condition, is given by

$$\begin{aligned} \Pr\{X(t) = n\} &= \frac{(at)^{n-1}}{(1 + at)^{n+1}},\ n = 1, 2, \ldots \\ &= \frac{at}{1 + at},\ n = 0. \end{aligned}$$

We have

$$\begin{aligned} E\{X(t)\} &= \sum_{n=0}^{\infty} n \Pr\{X(t) = n\} \\ &= \sum_{n=1}^{\infty} \frac{n(at)^{n-1}}{(1 + at)^{n+1}} \end{aligned}$$

$$= \frac{1}{(1+at)^2} \sum_{n=1}^{\infty} n \left(\frac{at}{1+at} \right)^{n-1} = 1$$

$$E\{X^2(t)\} = \sum_{n=2}^{\infty} \frac{n^2 (at)^{n-1}}{(1+at)^{n+1}}$$

$$= \frac{at}{(1+at)^3} \left\{ \sum_{n=2}^{\infty} n(n-1) \left(\frac{at}{1+at} \right)^{n-2} \right\}$$

$$+ \sum_{n=1}^{\infty} \frac{n(at)^{n-1}}{(1+at)^{n+1}}$$

$$= 2at + 1.$$

Thus,

$$\text{var}\,\{X(t)\} = 2at.$$

Here the first moment is constant but the second moment (and the variance) increases with t. The process $\{X(t)\}$ is not stationary.

EXERCISES

2.1. Let X_n, for n even, take values $+1$ and -1 each with probability $\frac{1}{2}$, and for n odd, take values $\sqrt{a}, -1/\sqrt{a}$, with probability $1/(a+1)$, $a/(a+1)$ respectively (a is a real number > -1 and $\neq 0, 1$). Further let X_n be independent. Show that $\{X_n\}$ is covariance stationary but not strictly stationary. (Compare this with Example 3 (a).)

2.2. Let $Y_n = a_0 X_n + a_1 X_{n-1}$ ($n = 1, 2, \ldots$), where a_0, a_1 are constants and X_n ($n = 0,1,2,\ldots$) are i.i.d. random variables with mean 0 and variance σ^2. Is $\{Y_n, n \geqslant 1\}$ covariance stationary ? Is it a Markov process ?

2.3. Let $X(t) = \sum_{r=1}^{k} (A_r \cos \theta_r t + B_r \sin \theta_r t)$,

where A_r, B_r are uncorrelated random variables with mean 0, variance σ^2 and θ_r are constants. Show that $\{X(t)\}$ is covariance stationary.

2.4. Let $X(t) = A_0 + A_1 t + A_2 t^2$, where A_i, $i = 0,1,2$ are uncorrelated random variables with mean 0 and variance 1. Find the mean value function and the covariance function of $\{X(t)\}$.

2.5. Find the covariance function of $\{Y_n\}$ given by

$$Y_n = a_0 X_n + a_1 X_{n-1} + \ldots + a_k X_{n-k},\ n = 1, 2, \ldots,$$

where a's are constants and X_n's are uncorrelated random variables.

2.6. Show that every stochastic process $\{X(t), t = 0,1,2\ldots\}$ with independent increments is a Markov process. Is the converse true ?

2.7. If $\{X(t), t \geqslant 0\}$ with $X(0) = 0$ is a process having independent increments, then

$$\begin{aligned}\operatorname{cov}\{X(s), X(t)\} &= \operatorname{var}\{X(s)\}, \text{ for } 0 \leqslant s \leqslant t\\ &= \operatorname{var}\{X(\min(s, t))\}, \text{ for any } s, t \ (t > 0, s > 0).\end{aligned}$$

(Such a process is not covariance stationary.)

In exercises 8-9, the distribution of $N(t)$ (the number of events that occur in $(0, t]$ is assumed to be Poisson with mean at, i.e. $\{N(t), t \geqslant 0\}$ is assumed to be a Poisson Process.

2.8. Poisson increment process. The process $\{X(t), t \geqslant 0\}$ defined by

$$X(t) = \frac{N(t+\varepsilon) - N(t)}{\varepsilon},$$

where $\varepsilon > 0$ is a given constant, is known as Poisson increment process. Show that the process is covariance stationary having

$$E\{X(t)\} = a$$

and

$$\begin{aligned}C(s, t) &= 0, \qquad |t-s| > \varepsilon\\ &= \frac{a}{\varepsilon} - \frac{a|t-s|}{\varepsilon^2}, \ |t-s| < \varepsilon.\end{aligned}$$

2.9.(a) Semi-random telegraph signal: The process $\{X(t), t \geqslant 0\}$ defined by

$$X(t) = (-1)^{N(t)}$$

is called a semi-random telegraph signal. Show that the process is evolutionary having

$$E\{X(t)\} = \exp(-2at)$$

and

$$C(s, t) = \exp\{-2a|t-s|\} - \exp\{-2a(t+s)\}.$$

The process $\{X(t)\}$ is a two-valued process with two possible values 1 and -1. It is a *one-minus-one* process.

(b) Random telegraph signal : The process $\{Y(t), t \geqslant 0\}$ defined by

$$Y(t) = \alpha X(t),$$

where α is a random variable which assumes the values -1 and 1 with equal probability and which is independent of $X(t)$ (defined in (a)), is called a random telegraph signal. Show that the process is covariance stationary having

$$E\{Y(t)\} = 0$$

and

$$C(s, t) = \exp\{-2a|t-s|\}.$$

(Random telegraph signals are useful in the construction of random signal generators.)

2.10. Semi-binary random transmission : Assume that a fair coin is tossed every T units of time. The process $\{Z(t), t \geqslant 0\}$ such that $Z(t)$ takes the values 1 or -1 according as nth toss results in a head or tail respectively, where $(n-1)T < t < nT$, is called semi-random binary transmission. Show that the process is covariance stationary having

$$E\{Z(t)\} = 0$$

and

$$C(s,t) = 1, \text{ if } s, t \text{ are in the same tossing interval}$$

$$(n-1)\,T < s,t < nT$$

$$= 0, \text{ otherwise.}$$

REFERENCES

J.L. Doob, *Stochastic Processes*, Wiley, New York (1953).

S. Karlin, *A First Course in Stochastic Processes*, Academic Press, New York (1966).

J. Neyman, "Indeterminism in Science and new demands for statisticians," *J. Am. Stat. Ass.*. 55, 625–639 (1960).

N.U. Prabhu, *Stochastic Processes*, Macmillan, New York (1965).

CHAPTER 3

MARKOV CHAINS

3.1 Definition and Examples

Consider a simple coin tossing experiment repeated for a number of times. The possible outcomes at each trial are two : head with probability, say, p and tail with probability q, $p + q = 1$. Let us denote head by 1 and tail by 0 and the random variable denoting the result of the nth toss by X_n. Then for $n = 1, 2, 3, \ldots$,

$$\Pr\{X_n = 1\} = p, \Pr\{X_n = 0\} = q.$$

Thus we have a sequence of random variables $X_1, X_2, \ldots$. The trials are independent and the result of the nth trial does not depend in any way on the previous trials numbered $1, 2, \ldots, (n-1)$.

Consider now the random variable given by the partial sum $S_n = X_1 + \ldots + X_n$. The sum S_n gives the accumulated number of heads in the first n trials and its possible values are $0, 1, \ldots, n$.

We have $S_{n+1} = S_n + X_{n+1}$. Given that $S_n = j$ $(j = 0, 1, \ldots n)$, the r.v. S_{n+1} can assume only two possible values : $S_{n+1} = j$ with probability q and $S_{n+1} = j + 1$ with probability p; these probabilities are not at all affected by the values of the variables $S_1, \ldots, S_{n-1}$. Thus

$$\Pr\{S_{n+1} = j + 1 \mid S_n = j\} = p$$
$$\Pr\{S_{n+1} = j \mid S_n = j\} = q.$$

We have here an example of a Markov* chain, a case of simple dependence that the outcome of $(n + 1)$st trial depends directly on that of nth trial and *only* on it. The conditional probability of S_{n+1} given S_n depends on the value of S_n and the manner in which the value of S_n was reached is of no consequence.

Example 1(a). Polya's urn model : An urn contains b black and r red balls. A ball is drawn at random and is replaced after the drawing, i.e. drawing is with replacement. The outcome at the nth drawing is either a black ball or a red ball. Let the random variable X_n be defined as

$X_n = 1$, if nth drawing results in a black ball, and

$X_n = 0$, if it results in a red ball.

---

* A.A. Markov (1856–1922).

There are two possible outcomes of X_n with

$$\Pr\{X_n = 1\} = b/(b+r),\ \Pr\{X_n = 0\} = r/(b+r).$$

We have

$$\Pr\{X_1 = j, \ldots, X_n = k\} = \Pr\{X_1 = j\} \ldots \Pr\{X_n = k\},\ j, k = 0, 1,$$

because of independence of $X_1, \ldots, X_n$.

Polya's urn model is such that after each drawing not only that the ball drawn is replaced but $c(>0)$ balls of the colour drawn (at that drawing) are added to the urn so that the number of balls of the colour drawn increases, while the number of balls of the other colour remains unchanged as at the drawing, i.e. if the nth drawing results in a black ball, then the number of black balls for $(n+1)$st drawing will be c more than the number of black balls at the nth drawing, while the number of red balls will be the same for both the drawings. We have

$$\begin{aligned}
\Pr\{X_2 = 1\} &= \Pr\{X_2 = 1, X_1 = 0\} + \Pr\{X_2 = 1, X_1 = 1\} \\
&= \Pr\{X_2 = 1 \mid X_1 = 0\} \Pr\{X_1 = 0\} + \Pr\{X_2 = 1 \mid \{X_1 = 1\} \Pr\{X_1 = 1\} \\
&= \frac{b}{b+r+c} \cdot \frac{r}{b+r} + \frac{b+c}{b+r+c} \cdot \frac{b}{b+r} \\
&= \frac{b}{b+r}.
\end{aligned}$$

It can be shown that $\Pr\{X_n = 1\} = b/(b+r)$, $\Pr\{X_n = 0\} = r/(b+r)$ for all n. Again,

$$\begin{aligned}
&\Pr\{X_3 = 1, X_2 = 1\} \\
&\quad = \Pr\{X_3 = 1, X_2 = 1, X_1 = 0\} + \Pr\{X_3 = 1, X_2 = 1, X_1 = 1\} \\
&\quad = \Pr\{X_3 = 1 \mid X_2 = 1, X_1 = 0\} \Pr\{X_2 = 1 \mid X_1 = 0\} \Pr\{X_1 = 0\} \\
&\qquad + \Pr\{X_3 = 1 \mid X_2 = 1, X_1 = 1\} \Pr\{X_2 = 1 \mid X_1 = 1\} \Pr\{X_1 = 1\} \\
&\quad = \frac{b+c}{b+r+2c} \cdot \frac{b}{b+r+c} \cdot \frac{r}{b+r} + \frac{b+2c}{b+r+2c} \cdot \frac{b+c}{b+r+c} \cdot \frac{b}{b+r} \\
&\quad = \frac{b(b+c)}{(b+r)(b+r+c)}.
\end{aligned}$$

Hence

$$\begin{aligned}
\Pr\{X_3 = 1 \mid X_2 = 1\} &= \frac{\Pr\{X_3 = 1, X_2 = 1\}}{\Pr\{X_2 = 1\}} \\
&= \frac{b+c}{b+r+c};
\end{aligned}$$

but

$$\Pr\{X_3 = 1 \mid X_2 = 1, X_1 = 1\} = \frac{b+2c}{b+r+2c}$$

and

$$\Pr\{X_3 = 1 \mid X_2 = 1, X_1 = 0\} = \frac{b+c}{b+r+2c}.$$

Hence

$$\Pr\{X_3 = 1 \mid X_2 = 1\} \neq \Pr\{X_3 = 1 \mid X_2 = 1, X_1 = k\}, k = 0, 1.$$

Here the conditional probability of X_3 given X_2 is not equal to that of X_3 given X_2 and X_1.

Now, consider the random variable Y_n which gives the number of black balls at the time of nth drawing. Y_1 takes only one value b, Y_2 takes values b and $b + c$, Y_3 values $b, b + c$ and $b + 2c$ etc. In general, Y_n can assume values $b, b + c, \ldots, b + (n-1)c$. Given that Y_{n-1} assumes one of the possible values $b, b + c, \ldots, b + (n - 2)c$, say $i = b + kc$, $k = 0, 1, \ldots, n - 2$, then Y_n assumes the value $j = i$ with probability a_1 and the value $j = i + c$ with probability b_1, where

$$a_1 = \Pr\{X_{n-1} = 0\} = \frac{r}{b + r}$$

$$b_1 = \Pr\{X_{n-1} = 1\} = \frac{b}{b + r}.$$

Let $\{X_n, n = 0, 1, 2, \ldots\}$ be a sequence of random variables (or a discrete parameter stochastic process). Let the possible outcomes of X_n be E_j $(j = 0, 1, 2, \ldots)$, where the number of E_j may be finite (say, m) or denumerable. To the event E_j let us associate the value $j(=0, 1, 2, \ldots)$ so that the possible values of X_n constitute a set $N = \{0, 1, 2 \ldots\}$, and that the process has the state space N. Unless otherwise stated, by state space of a Markov chain, we shall imply *discrete state space* (having finite or a countably infinite number of elements); it could be N or some other subset of the set of integers $I = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$.

Definition. The stochastic process $\{X_n, n = 0, 1, 2, \ldots\}$ is called a Markov chain, if, for $j, k, j_1, \ldots j_{n-1} \in N$ (or any subset of I),

$$\begin{aligned}\Pr\{X_n = k \mid X_{n-1} = j, X_{n-2} = j_1, \ldots, X_0 = j_{n-1}\}\\ = \Pr\{X_n = k \mid X_{n-1} = j\} = p_{jk} \text{ (say)},\end{aligned} \tag{1.1}$$

whenever the first member is defined.

The outcomes E_j (or the values j) are called the states of the Markov chain; if X_n has the outcome E_j (i.e. $X_n = j$), the process is said to be at state E_j **(or simply at state** j) at nth trial. To state j (or outcome E_j) there is no longer a fixed probability $\Pr\{X_n = j\}$ but to a pair of states (j, k) at the two successive trials (say, nth and $(n + 1)$st trials) there is a conditional probability p_{jk}. It is the probability of transition from the state j at nth trial to the state k at $(n + 1)$st trial. The transition probabilities p_{jk} are basic to the study of the structure of the Markov chain.

The transition probability may or may not be independent of n. If the transition probability p_{jk} is independent of n, the Markov chain is said to be *homogeneous* (or to have *stationary transition probabilities*). If it is dependent on n (in which case it is denoted by $({}^{n}p_{ij})$, the chain is said to be non-homogeneous. *Here we shall confine to homogeneous chains.*

The transition probability p_{jk} refers to the states (j, k) at two *successive* trials (say nth and $(n+1)$st trial); the transition is one-step and p_{jk} is called one-step (or unit step) transition probability. In the more general case, we are concerned with the pair of states (j, k) at two non-successive trials, say, state j at the nth trial and state k at the $(n+m)$th trial. The corresponding transition probability is then called m-step transition probability and is denoted by $p_{jk}^{(m)}$, i.e.

$$p_{jk}^{(m)} = \Pr\{X_{n+m} = k \mid X_n = j\}.$$

This will be considered in Sec. 3.2 below.

Transition Matrix: The transition probabilities p_{jk} satisfy

$$p_{jk} \geqslant 0, \quad \sum_k p_{jk} = 1 \text{ for all } j. \tag{1.2}$$

These probabilities may be written in the matrix form

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{1.3}$$

This is called the *transition probability matrix* or *matrix of transition probabilities* (t.p.m.) of the Markov chain. P is a *stochastic* or *Markov matrix*, i.e. a square matrix with non-negative elements and unit row sums.

It may be seen that the probability distribution of $X_r, X_{r+1}, \ldots, X_{r+n}$ can be computed in terms of the transition probabilities p_{jk} and the initial distribution of X_r. Suppose, for simplicity, that $r = 0$, then

$$\begin{aligned}
&\Pr\{E_a, E_b, \ldots, E_i, E_j, E_k\} \\
&\quad = \Pr\{X_0 = a, X_1 = b, \ldots, X_{n-2} = i, X_{n-1} = j, X_n = k\} \\
&\quad = \Pr\{X_n = k \mid X_{n-1} = j, \ldots, X_0 = a\} \Pr\{X_{n-1} = j, \ldots, X_0 = a\} \\
&\quad = \Pr\{X_n = k \mid X_{n-1} = j\} \Pr\{X_{n-1} = j \mid X_{n-2} = i\} \\
&\qquad\qquad . \Pr\{X_{n-2} = i, \ldots, X_0 = a\} \\
&\quad = \Pr\{X_n = k \mid X_{n-1} = j\} \Pr\{X_{n-1} = j \mid X_{n-2} = i\} \ldots \\
&\qquad\qquad \ldots \Pr\{X_1 = b \mid X_0 = a\} \Pr\{X_0 = a\} \\
&\quad = (\Pr\{X_0 = a\})\, p_{ab} \ldots p_{ij}\, p_{jk}.
\end{aligned} \tag{1.4}$$

Thus,

$$\begin{aligned}
&\Pr\{X_r = a, X_{r+1} = b, \ldots, X_{r+n-2} = i, X_{r+n-1} = j, X_{r+n} = k\} \\
&\quad = (\Pr\{X_r = a\})\, p_{ab} \ldots p_{ij}\, p_{jk}.
\end{aligned} \tag{1.4a}$$

More generally, if we denote by E_{j_r} the outcome of the rth trial then

$$\Pr\{E_{j_0}, E_{j_1}, \ldots, E_{j_n}\} = (\Pr\{E_{j_0}\})\, p_{j_0 j_1} p_{j_1 j_2} \cdots p_{j_{(n-1)} j_n}$$

The matrix of transition probabilities together with the initial distribution completely specifies a Markov chain $\{X_n, n = 0, 1, 2, \ldots\}$.

We state (without proof) the general existence theorem of Markov chains.

Given the set N and the sequence of stochastic matrices $({}^{(n)}p_{jk}) = {}^{(n)}P$, there exists a Markov chain $\{X_n\}$ with state space N and transition probability matrices, ${}^{(n)}P$. (For proof, see Iosifescu & Tautu, *Stochastic Processes*—I, Springer-Verlag (1973), Chung (1967).)

Order of a Markov chain

Definition. A Markov chain $\{X_n\}$ is said to be of order $s(s = 1, 2, 3, \ldots)$, if, for all n,

$$\Pr\{X_n = k \mid X_{n-1} = j, X_{n-2} = j_1, \ldots, X_{n-s} = j_{s-1}, \ldots\}$$
$$= \Pr\{X_n = k \mid X_{n-1} = j, \ldots, X_{n-s} = j_{s-1}\}. \quad (1.5)$$

A Markov chain $\{X_n\}$ is said to be of order one (or simply a Markov chain) if

$$\Pr\{X_n = k \mid X_{n-1} = j, X_{n-2} = j_1, \ldots\} = \Pr\{X_n = k \mid X_{n-1} = j\}$$
$$= p_{jk} \quad (1.6)$$

Unless explicitly stated otherwise, we shall mean by Markov chain, a *chain of order one*, to which we shall mostly confine ourselves here. A chain is said to be of order zero if $p_{jk} = p_k$ for all j. This implies independence of X_n and X_{n-1}.

Denote the state that a day is rainy by 1 and that a day is not rainy by 0.

Let (1.5) hold for $s = 2$ and let

$p_{ijk} = \Pr\{$actual day is in state $k \mid$ the preceding day was in state j, the day before the preceding was in state $i\}$, $i, j, k = 0, 1$.

We have then a Markov chain of order two. Note that the matrix (p_{ijk}) is not a stochastic matrix.

Let (1.5) hold for $s = 1$ and let

$p_{jk} = \Pr\{$actual day is in state $k \mid$ preceding day was in state $j\}$.

We then have a Markov chain (i.e. a chain of order one) with t.p.m. (p_{jk}), $j, k = 0, 1$.

Example 1(b). Let $\{X_n, n \geqslant 0\}$ be a Markov chain with three states 0, 1, 2 and with transition matrix

$$\begin{pmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 3/4 & 1/4 \end{pmatrix} \quad (1.7)$$

and the initial distribution $\Pr\{X_0 = i\} = \frac{1}{3}$, $i = 0, 1, 2$.

We have

$$\Pr\{X_1 = 1 \mid X_0 = 2\} = \tfrac{3}{4}$$

$$\Pr\{X_2 = 2 \mid X_1 = 1\} = \tfrac{1}{4}$$

$$\begin{aligned}\Pr\{X_2 = 2, X_1 = 1 \mid X_0 = 2\} \\ &= \Pr\{X_2 = 2 \mid X_1 = 1\}\Pr\{X_1 = 1 \mid X_0 = 2\} = \tfrac{1}{4}\cdot\tfrac{3}{4} = \tfrac{3}{16}\end{aligned}$$

$$\begin{aligned}\Pr\{X_2 = 2, X_1 = 1, X_0 = 2\} \\ &= \Pr\{X_2 = 2, X_1 = 1 \mid X_0 = 2\}\Pr\{X_0 = 2\} = \tfrac{3}{16}\cdot\tfrac{1}{3} = \tfrac{1}{16}\end{aligned}$$

$$\begin{aligned}\Pr\{X_3 = 1, X_2 = 2, X_1 = 1, X_0 = 2\} \\ &= \Pr\{X_3 = 1 \mid X_2 = 2, X_1 = 1, X_0 = 2\}.\Pr\{X_2 = 2, \\ &\qquad X_1 = 1, X_0 = 2\} \\ &= \Pr\{X_3 = 1 \mid X_2 = 2\}\left(\tfrac{1}{16}\right) = \tfrac{3}{4}\cdot\tfrac{1}{16} = \tfrac{3}{64}.\end{aligned}$$

Example 1(c). A particle performs a random walk with absorbing barriers, say, at 0 and 4. Whenever it is at any position r $(0 < r < 4)$, it moves to $r + 1$ with probability p or to $(r - 1)$ with probability q, $p + q = 1$. But as soon as it reaches 0 or 4 it remains there itself. Let X_n be the position of the particle after n moves. The different states of X_n are the different positions of the particle. $\{X_n\}$ is a Markov chain whose unit-step transition probabilities are given by

$$\left.\begin{aligned}\Pr\{X_n = r + 1 \mid X_{n-1} = r\} &= p \\ \Pr\{X_n = r - 1 \mid X_{n-1} = r\} &= q\end{aligned}\right\} \qquad 0 < r < 4$$

$$\Pr\{X_n = 0 \mid X_{n-1} = 0\} = 1,$$

and

$$\Pr\{X_n = 4 \mid X_{n-1} = 4\} = 1.$$

The transition matrix is given by

$$\begin{array}{cc} & \text{States of } X_n \\ \text{States of } X_{n-1} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \end{array}$$

Example 1(d). Suppose that a coin with probability p for a head is tossed indefinitely. Let X_n, the outcome of the nth trial, be k, where k $(= 0, 1, \ldots, n)$ denotes that there is a run of k successes, i.e. the length of the uninterrupted block of heads is k. $\{X_n\}$ constitutes a Markov chain, with unit-step transition probabilities

$$\begin{aligned} p_{jk} = \Pr\{X_n = k \mid X_{n-1} = j\} &= p, \quad k = j + 1 \\ &= q, \quad k = 0 \\ &= 0, \quad \text{otherwise.} \end{aligned}$$

The transition matrix is given by

$$\begin{array}{ccl}
 & \nearrow & \text{States of } X_n \\
\text{States of } X_{n-1} & & \begin{array}{ccccccc} 0 & 1 & 2 & \cdots & k & k+1 & \cdots \end{array} \\
 & \begin{array}{c} 0 \\ 1 \\ 2 \\ \cdot \\ \cdot \\ \cdot \\ k \\ \cdot \\ \cdot \\ \cdot \end{array} & \left[\begin{array}{ccccccc}
q & p & 0 & \cdots & 0 & 0 & \cdots \\
q & 0 & p & \cdots & 0 & 0 & \cdots \\
q & 0 & 0 & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
q & 0 & 0 & \cdots & 0 & p & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots
\end{array}\right]
\end{array}$$

Example 1(e). Partial sum of independent random variables: Consider a series of coin tossing experiments, where the outcomes of nth trial are denoted by 1 (for a head) and 0 (for a tail). Let X_n be the random variable denoting the outcome of nth trial and $S_n = X_1 + \ldots + X_n$ be the nth partial sum. The possible values of S_n are $0, 1, \ldots, n$, i.e. the states of S_n are r, $r = 0, 1, \ldots n$. $\{S_n\}$ is a Markov chain with transition matrix

$$\begin{array}{ccl}
 & \nearrow & \text{States of } X_n \\
\text{States of } X_{n-1} & & \begin{array}{ccccccc} 1 & 2 & 3 & \cdots & r & r+1 & \cdots \end{array} \\
 & \begin{array}{c} 1 \\ 2 \\ 3 \\ \cdot \\ \cdot \\ \cdot \\ r \\ \cdot \\ \cdot \\ \cdot \end{array} & \left[\begin{array}{ccccccc}
q & p & 0 & \cdots & 0 & 0 & \cdots \\
0 & q & p & \cdots & 0 & 0 & \cdots \\
0 & 0 & q & \cdots & 0 & 0 & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
0 & 0 & 0 & \cdots & q & p & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdots
\end{array}\right]
\end{array}$$

A Markov chain $\{X_n, n \geqslant 0\}$ with k states, where k is finite, is said to be a finite Markov chain. The transition matrix P is in this case a square matrix with k rows and k columns. Examples 1(b) and 1(c) deal with finite Markov chains.

The number of states could however be infinite. If the possible values of X_n are $\ldots, -2, -1, 0, 1, 2, \ldots$, then the Markov chain is said to be *denumerably infinite* or *denumerable*. Examples 1(d) and 1(e) are of denumerable Markov chains.

3.2 Higher Transition Probabilities

Chapman-Kolmogorov equation: We have so far considered unit-step or one-step transition probabilities, the probability of X_n given X_{n-1}, i.e. the probability of the outcome at the nth step or trial given the outcome at the previous step. p_{jk} gives the probability of unit-step transition from the state j at a trial to the state k at the next following trial. The m-step transition probability is given by

$$\Pr\{X_{m+n} = k \mid X_n = j\} = p_{jk}^{(m)}; \tag{2.1}$$

$p_{jk}^{(m)}$ gives the probability that from the state j at nth trial, the state k is reached at $(m+n)$th trial in m steps, i.e. the probability of transition from the state j to the state k in exactly m steps. The number n does not occur in the r.h.s. of the relation (2.1) and the chain is *homogeneous*. The one-step transition probabilities $p_{jk}^{(1)}$ are denoted by p_{jk} for simplicity. Consider

$$p_{jk}^{(2)} = \Pr\{X_{n+2} = k \mid X_n = j\}. \tag{2.2}$$

The state k can be reached from the state j in two steps through some intermediate state r. Consider a fixed value of r; we have

$$\begin{aligned}\Pr\{X_{n+2} = k,\ X_{n+1} = r \mid X_n = j\} \\ &= \Pr\{X_{n+2} = k \mid X_{n+1} = r, X_n = j\} \Pr\{X_{n+1} = r \mid X_n = j\} \\ &= p_{rk}^{(1)} p_{jr}^{(1)} = p_{jr}\, p_{rk}.\end{aligned}$$

Since these intermediate states r for $r = 1, 2, \ldots$ are mutually exclusive, we have

$$\begin{aligned}p_{jk}^{(2)} = \Pr\{X_{n+2} = k \mid X_n = j\} &= \sum_r \Pr\{X_{n+2} = k,\ X_{n+1} = r \mid X_n = j\} \\ &= \sum_r p_{jr}\, p_{rk}\end{aligned} \tag{2.3}$$

(summing over all the intermediate states).

By induction, we have

$$\begin{aligned}p_{jk}^{(m+1)} &= \Pr\{X_{n+m+1} = k \mid X_n = j\} \\ &= \sum_r \Pr\{X_{n+m+1} = k \mid X_{n+m} = r\} \Pr\{X_{n+m} = r \mid X_n = j\} \\ &= \sum_r p_{rk} p_{jr}^{(m)}.\end{aligned}$$

Similarly, we get

$$p_{jk}^{(m+1)} = \sum_r p_{jr}\, p_{rk}^{(m)}.$$

In general, we have

$$p_{jk}^{(m+n)} = \sum_r p_{rk}^{(n)} p_{jr}^{(m)} = \sum_r p_{jr}^{(n)}\, p_{rk}^{(m)}. \tag{2.4}$$

This equation is a special case of *Chapman-Kolmogorov* equation, which is satisfied by the transition probabilities of a Markov chain.

We can put the results in terms of transition matrices as follows. Let $P = (p_{jk})$ denote the transition matrix of the unit-step transitions and $P^{(m)} = (p_{jk}^{(m)})$ denote the transition matrix of the m-step transitions. For $m = 2$, we have the matrix $P^{(2)}$ whose elements are given by (2.3). It follows that the elements of $P^{(2)}$ are the elements of the matrix obtained by multiplying the matrix P by itself, i.e.

$$P^{(2)} = P \cdot P = P^2.$$

Similarly,

$$P^{(m+1)} = P^m \cdot P = P \cdot P^m$$

and

$$P^{(m+n)} = P^m \cdot P^n = P^n \cdot P^m. \tag{2.5}$$

It should be noted that there exist non-Markovian chains whose transition probabilities satisfy Chapman-Kolmogorov equation (for example, see Feller I, p. 423, Parzen p. 203).

Example 2(a). Consider the Markov chain of Example 1(b). The two-step transition matrix is given by

$$\begin{pmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 3/4 & 1/4 \end{pmatrix} \begin{pmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 3/4 & 1/4 \end{pmatrix} = \begin{pmatrix} 5/8 & 5/16 & 1/16 \\ 5/16 & 1/2 & 3/16 \\ 3/16 & 9/16 & 1/4 \end{pmatrix}$$

Hence $p_{01}^{(2)} = \Pr\{X_{n+2} = 1 \mid X_n = 0\} = 5/16$ for $n \geqslant 0$.
Thus

$$\Pr\{X_2 = 1 \mid X_0 = 0\} = 5/16,$$

and

$$\Pr\{X_2 = 1, X_0 = 0\} = \Pr\{X_2 = 1 \mid X_0 = 0\} \Pr\{X_0 = 0\}$$
$$= (5/16).(1/3) = 5/48.$$

Example 2(b). Two-state Markov chain. Suppose that the probability of a dry day (state 0) following a rainy day (state 1) is 1/3 and that the probability of a rainy day following a dry day is 1/2. We have a two-state Markov chain such that $p_{10} = 1/3$ and $p_{01} = 1/2$ and t.p.m.

$$P = \begin{array}{c|cc} & 0 & 1 \\ 0 & 1/2 & 1/2 \\ 1 & 1/3 & 2/3 \end{array}$$

We have

$$P^2 = \begin{pmatrix} 5/12 & 7/12 \\ 7/18 & 11/18 \end{pmatrix}, \quad P^4 = \begin{pmatrix} 173/432 & 259/432 \\ 259/648 & 389/648 \end{pmatrix}.$$

Given that May 1 is a dry day, the probability that May 3 is a dry day is 5/12, and that May 5 is a dry day is 173/432. For calculation of powers of P, refer to Sec. 3.4.

Example 2(c). Consider a communication system which transmits the two digits 0 and 1 through several stages. Let $\{X_n, n \geqslant 1\}$ be the digit leaving the nth stage of system and X_0 be the digit entering the first stage (leaving the 0th stage). At each stage there is a constant probability q that the digit which enters will be transmitted unchanged (i.e. the digit will remain unchanged when it leaves), and probability p otherwise (i.e. the digit changes when it leaves), $p + q = 1$.

Here $\{X_n, n \geqslant 0\}$ is a homogeneous two-state Markov chain with unit-step transition matrix

$$P = \begin{pmatrix} q & p \\ p & q \end{pmatrix}.$$

It can be shown (by mathematical induction or otherwise) that

$$P^m = \begin{pmatrix} \frac{1}{2} + \frac{1}{2}(q-p)^m & \frac{1}{2} - \frac{1}{2}(q-p)^m \\ \frac{1}{2} - \frac{1}{2}(q-p)^m & \frac{1}{2} + \frac{1}{2}(q-p)^m \end{pmatrix}.$$

Here

$$p_{00}^{(m)} = p_{11}^{(m)} = \tfrac{1}{2} + \tfrac{1}{2}(q-p)^m$$

and

$$p_{01}^{(m)} = p_{10}^{(m)} = \tfrac{1}{2} - \tfrac{1}{2}(q-p)^m.$$

and as $m \to \infty$,

$$\lim p_{00}^{(m)} = \lim p_{01}^{(m)} = \lim p_{10}^{(m)} = \lim p_{11}^{(m)} \to \tfrac{1}{2}.$$

Suppose that the initial distribution is given by

$$\Pr\{X_0 = 0\} = a \text{ and } \Pr\{X_0 = 1\} = b = 1 - a.$$

Then we have

$$\begin{aligned} \Pr\{X_m = 0, X_0 = 0\} &= \Pr\{X_m = 0 \mid X_0 = 0\} \Pr\{X_0 = 0\} \\ &= a p_{00}^{(m)} \end{aligned}$$

and

$$\Pr\{X_m = 0, X_0 = 1\} = b p_{10}^{(m)}.$$

The probability that the digit entering the first stage is 0 given that the digit leaving the mth stage is 0 can be evaluated by applying Bayes' rule. We have

$$\begin{aligned} &\Pr\{X_0 = 0 \mid X_m = 0\} \\ &= \frac{\Pr\{X_m = 0 \mid X_0 = 0\} \Pr\{X_0 = 0\}}{\Pr\{X_m = 0 \mid X_0 = 0\} \Pr\{X_0 = 0\} + \Pr\{X_m = 0 \mid X_0 = 1\} \Pr\{X_0 = 1\}} \\ &= \frac{a p_{00}^{(m)}}{a p_{00}^{(m)} + b p_{10}^{(m)}} \\ &= \frac{a\{\frac{1}{2} + \frac{1}{2}(q-p)^m\}}{a\{\frac{1}{2} + \frac{1}{2}(q-p)^m\} + b\{\frac{1}{2} - \frac{1}{2}(q-p)^m\}} \\ &= \frac{a\{1 + (q-p)^m\}}{1 + (a-b)(q-p)^m}. \end{aligned}$$

3.3 Classification of States and Chains

The states j, $j = 0,1,2,\ldots$ of a Markov chain $\{X_n, n \geqslant 0\}$ can often be classified in a distinctive manner according to some fundamental properties of the system. By means of such classification it is possible to identify certain types of chains.

If the probability $p_{jk}^{(n)}$ is non-zero for some $n \geqslant 1$, then we say that the state k can be reached from the state j. If every state can be reached from every other state (in *any* number of transitions) the chain is said to be *irreducible*. The transition matrix is then irreducible.

If C is a set of states such that no state outside C can be reached from any state in C, then the set C is said to be *closed*. If C is a closed set and $j \in C$ while $k \notin C$, then $p_{jk} = 0$. It can be seen that $p_{jk}^{(2)} = 0$, since one of the factors in each term of the sum given by (2.3) will be zero. More generally, $p_{jk}^{(n)} = 0$ for $n \geqslant 1$.

A closed set may contain one or more states. If a closed set C contains only one state j, then the state is called an *absorbing state*. j is absorbing iff $p_{jj} = 1$ and then $p_{jk} = 0$ for $k \neq j$. The set of all states of a Markov chain will be a closed set. It follows that if a Markov chain does not contain any other closed set (with the exception of the set of all states), then the chain will be *irreducible*, i.e. an irreducible chain does not contain any proper closed subset. Chains which are not irreducible are said to be non-irreducible or reducible, or decomposable. Every Markov chain will contain at least one closed set, i.e. the set of all states. If the number of closed sets is two or more, the chain is reducible.

In Example 1(c), the states 0 and 4 are absorbing states. In Example 2(c), for $p \neq 0, 1$ (and $q \neq 0, 1$) both the states can be reached from either of the states, i.e. the Markov chain is irreducible.

We now proceed to obtain a classification of the states of a Markov chain.

Suppose that a system starts with the state j. Let $f_{jk}^{(n)}$ be the probability that it reaches the state k for the *first time* at the nth step (or after n transitions) and let $p_{jk}^{(n)}$ be the probability that it reaches the state k (not necessarily for the first time) after n transitions. A relation can be established between $f_{jk}^{(n)}$ and $p_{jk}^{(n)}$ as follows :

The probability that starting with j, the state k is reached for the first time at the rth step and again after that at the $(n-r)$th step is given by $f_{jk}^{(r)} p_{kk}^{(n-r)}$ for all $r \leqslant n$. These cases are mutually exclusive. Hence

$$p_{jk}^{(n)} = \sum_{r=0}^{n} f_{jk}^{(r)} p_{kk}^{(n-r)}, n \geqslant 1 \tag{3.1}$$

with

$$p_{kk}^{(0)} = 1, f_{jk}^{(0)} = 0, f_{jk}^{(1)} = p_{jk}$$

or,

$$p_{jk}^{(n)} = \sum_{r=1}^{n-1} f_{jk}^{(r)} p_{kk}^{(n-r)} + f_{jk}^{(n)}, \; n > 1. \tag{3.2}$$

Let F_{jk} denote the probability that starting with the state j the system will ever reach the state k. Clearly

$$F_{jk} = \sum_{n=1}^{\infty} f_{jk}^{(n)}. \tag{3.3}$$

The mean (*first-passage*) time from the state j to the state k is given by

$$\mu_{jk} = \sum_{n=1}^{\infty} n f_{jk}^{(n)}. \tag{3.4}$$

We have to consider two cases, $F_{jk} = 1$ and $F_{jk} < 1$.

When $F_{jk} = 1$, it is certain that the system starting with the state j will reach the state k; in this case $\{f_{jk}^{(n)}, n = 1, 2, \ldots\}$ is a proper probability distribution and this gives the *first passage distribution* for k (the system starting with j). In particular, when $k = j$, $\{f_{jj}^{(n)}, n = 1, 2 \ldots\}$ will represent the distribution of the *recurrence times* of j; and $F_{jj} = 1$ will imply that the return to the state j is certain. In this case

$$\mu_{jj} = \sum_{n=1}^{\infty} n f_{jj}^{(n)} \tag{3.5}$$

is known as the *mean recurrence time* for the state j.

Thus, two questions concerning state j arise: first, whether the return to the state j is certain and secondly, when this happens, whether the mean recurrence time μ_{jj} is finite.

Definitions. The state j is said to be *persistent** if $F_{jj} = 1$ (i.e. return to the state j is certain) and *transient* if $F_{jj} < 1$ (i.e. return to the state j is uncertain). A persistent state j is said to be null, if $\mu_{jj} = \infty$, i.e. if the mean recurrence time is infinite, and is said to be non-null if $\mu_{jj} < \infty$. A state j is said to be *periodic* with period $t (> 1)$ if return to the state is possible only at t, $2t$, $3t, \ldots$ steps, where t is the greatest integer with this property. In this case $p_{jj}^{(n)} = 0$ unless n is an integral multiple of t. In other words, if t is the greatest common divisor of all integers $n (\geqslant 1)$ for which $p_{jj}^{(n)} > 0$, then j is said to be periodic with period t. The state j is said to be *aperiodic* (or non-periodic) if no such $t (> 1)$ exists, e.g., $t = 1$ will correspond to the aperiodic case. A persistent, non-null and aperiodic state (of a Markov chain) is said to be *ergodic*. A Markov chain all of whose states are ergodic is said to be *an ergodic chain*. We consider below some important results concerning classification of states.

Theorem 3.1. The state j is persistent or transient according as

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} = \infty \text{ or } < \infty. \tag{3.6}$$

Proof: Let

$$P_{jj}(s) = \sum_{n=0}^{\infty} p_{jj}^{(n)} s^n = 1 + \sum_{n=1}^{\infty} p_{jj}^{(n)} s^n, \ |s| < 1$$

---

*The word *recurrent* is also used by some authors; we shall however use the word *persistent*.

and

$$F_{jj}(s) = \sum_{n=0}^{\infty} f_{jj}^{(n)} s^n = \sum_{n=1}^{\infty} f_{jj}^{(n)} s^n, \ |s| < 1$$

be the generating functions of the sequences $\{p_{jj}^{(n)}\}$ and $\{f_{jj}^{(n)}\}$ respectively. We have from (3.1)

$$p_{jj}^{(n)} = \sum_{r=0}^{n} f_{jj}^{(r)} p_{jj}^{(n-r)}. \tag{3.7}$$

Multiplying both sides of (3.7) by s^n and adding for all $n \geqslant 1$ we get

$$P_{jj}(s) - 1 = F_{jj}(s)\, P_{jj}(s).$$

The r.h.s. of the above is immediately obtained by considering the fact that the r.h.s. of (3.7) is a convolution of $\{f_{jj}\}$ and $\{p_{jj}\}$ and that the generating function of the convolution is the product of the two generating functions. Thus we have

$$P_{jj}(s) = \frac{1}{1 - F_{jj}(s)}, \ |s| < 1 \tag{3 8}$$

Using Abel's lemma, we have as $s \to 1$,

$$\begin{aligned} P_{jj}(s) &\to \frac{1}{1 - F_{jj}}, \text{ when } F_{jj} < 1 \\ &\to \infty, \text{ when } F_{jj} = 1, \end{aligned} \tag{3.9}$$

which proves the theorem. ▲

Corollary: If j is transient then $p_{jj}^{(n)} \to 0$ as $n \to \infty$.
This follows at once from the fact that when state j is transient, the series $\sum_{n} p_{jj}^{(n)}$ converges.

We now state a lemma without proof (for proof of this important basic limit theorem of renewal theory, see Prabhu, Feller).

Lemma 3.1 Let $\{f_n\}$ be a sequence such that $f_0 = 0, f_n \geqslant 0, \Sigma f_n = 1$ and $t (\geqslant 1)$ be the greatest common divisor of those n for which $f_n > 0$.

Let $\{u_n\}$ be another sequence such that $u_0 = 1$ and $u_n = \sum_{r=1}^{n} f_r u_{n-r}$ $(n \geqslant 1)$.

Then

$$\lim_{n\to\infty} u_{nt} = t/\mu, \tag{3.10}$$

where $\mu = \sum_{n=1}^{\infty} n f_n$, the limit being zero when $\mu = \infty$;

and

$$\lim_{N\to\infty} u_N = 0,$$

whenever N is not divisible by t

Theorem 3.2. If state j is persistent, then as $n \to \infty$,

(i)
$$p_{jj}^{(nt)} \to t/\mu \tag{3.11}$$

when state j is periodic with period t:

(ii) $$p_{jj}^{(n)} \rightarrow 1/\mu_{jj} \tag{3.12}$$

when state j is aperiodic.

In case state j is persistent null (whether periodic or aperiodic),

$$p_{jj}^{(n)} \rightarrow 0, \qquad \text{as } n \rightarrow \infty. \tag{3.13}$$

Proof: Let state j be persistent; then $\mu_{jj} = \sum_n n f_{jj}^{(n)}$ is defined.

Since (3.7) holds, we may put

$$f_{jj}^{(n)} \text{ for } f_n, \; p_{jj}^{(n)} \text{ for } u_n, \; \text{ and } \; \mu_{jj} \text{ for } \mu$$

in the lemma 3.1 above. Applying the lemma, we get

$$p_{jj}^{(nt)} \rightarrow t/\mu_{jj}, \text{ as } n \rightarrow \infty,$$

when state j is periodic with period t.

When state j is aperiodic (i.e. $t = 1$), then

$$p_{jj}^{(n)} \rightarrow 1/\mu_{jj}, \text{ as } n \rightarrow \infty.$$

In case state j is persistent null, $\mu_{jj} = \infty$,
and

$$p_{jj}^{(n)} \rightarrow 0 \qquad \text{as } n \rightarrow \infty. \quad ▲$$

Note that a persistent state j is non-null or null according as $\lim_{n\to\infty} p_{jj}^{(n)}$ is positive or zero.

Theorem 3.3. If state k is either transient or persistent null, then for every j

$$p_{jk}^{(n)} \rightarrow 0 \qquad \text{as } n \rightarrow \infty \tag{3.14}$$

and if state k is aperiodic persistent non-null, then

$$p_{jk}^{(n)} \rightarrow \frac{F_{jk}}{\mu_{kk}} \text{ as } n \rightarrow \infty. \tag{3.15}$$

Proof: We have

$$p_{jk}^{(n)} = \sum_{r=1}^{n} f_{jk}^{(r)} \; p_{kk}^{(n-r)}$$

Let $n > m$, then

$$\begin{aligned} p_{jk}^{(n)} &= \sum_{r=1}^{m} f_{jk}^{(r)} p_{kk}^{(n-r)} + \sum_{r=m+1}^{n} f_{jk}^{(r)} p_{kk}^{(n-r)} \\ &\leqslant \sum_{r=1}^{m} f_{jk}^{(r)} p_{kk}^{(n-r)} + \sum_{r=m+1}^{n} f_{jk}^{(r)} \end{aligned} \tag{3.16}$$

Since state k is either transient or persistent null, $p_{kk}^{(n-r)} \rightarrow 0$ as $n \rightarrow \infty$ (Theorem 3.2).

Further, $$\sum_{m=1}^{\infty} f_{jk}^{(m)} < \infty \text{ and } \sum_{r=m+1}^{n} f_{jk}^{(r)} \rightarrow 0 \text{ as } n, m \rightarrow \infty.$$

Hence as $n \to \infty$

$$p_{jk}^{(n)} \to 0.$$

From (3.16)

$$p_{jk}^{(n)} - \sum_{r=1}^{m} f_{jk}^{(r)} p_{kk}^{(n-r)} \leqslant \sum_{r=m+1}^{n} f_{jk}^{(r)}. \tag{3.16a}$$

Since j is aperiodic, persistent and non-null, then by Theorem 3.2,

$$p_{kk}^{(n-r)} \to 1/\mu_{kk} \text{ as } n \to \infty.$$

Hence from (3.16a) we get, as n, $m \to \infty$

$$p_{jk}^{(n)} - \frac{\sum\limits_{r=1}^{m} f_{jk}^{(r)}}{0,\mu_{kk}} = 0$$

or

$$p_{jk}^{(n)} \to \frac{F_{jk}}{\mu_{kk}}. \quad \blacktriangle$$

Theorem 3.4. In an irreducible chain, all the states are of the same type. They are either all transient, all persistent null, or all persistent non-null. All the states are aperiodic or periodic and in the latter case they all have the same period.

Proof: Since the chain is irreducible, every state can be reached from every other state. If i, j are any two states, then i can be reached from j and j from i, i.e.

$$p_{ij}^{(N)} = a > 0 \quad \text{for some } N \geqslant 1$$

and

$$p_{ji}^{(M)} = b > 0 \quad \text{for some } M \geqslant 1.$$

We have $p_{jk}^{(n+m)} = p_{jk}^{(m+n)} = \sum_r p_{jr}^{(m)} p_{rk}^{(n)}$

$$\geqslant p_{jr}^{(m)} p_{rk}^{(n)} \quad \text{for each } r.$$

Hence

$$p_{ii}^{(n+N+M)} \geqslant p_{ij}^{(N)} p_{jj}^{(n)} p_{ji}^{(M)} = ab\, p_{jj}^{(n)} \tag{3.17}$$

and

$$p_{jj}^{(n+N+M)} \geqslant p_{ji}^{(M)} p_{ii}^{(n)} p_{ij}^{(N)} = ab\, p_{ii}^{(n)}. \tag{3.18}$$

From the above it is clear that the two series $\sum_n p_{ii}^{(n)}$, and $\sum_n p_{jj}^{(n)}$ converge or diverge together. Thus the two states i, j are either both transient or both persistent.

Suppose that i is persistent null, then $p_{ii}^{(n)} \to 0$ as $n \to \infty$; then from (3.17), $p_{jj}^{(n)} \to 0$ as $n \to \infty$, so that j is also persistent null, i.e. they are both persistent null.

Suppose that i is persistent (non-null) and has period t, then $p_{ii}^{(n)} > 0$ whenever n is a multiple of t. Now

$$p_{ii}^{(N+M)} \geqslant p_{ij}^{(N)} p_{ji}^{(M)} = ab > 0,$$

so that $(N+M)$ is a multiple of t. From (3.18)

$$p_{jj}^{(n+N+M)} \geqslant ab\, p_{ii}^{(n)} > 0.$$

Thus $(n+N+M)$ is a multiple of t and so t is the period of the state j also. ▲

Two states are said to belong to the same *class* iff each one of them can be reached from the other. A property defined for all states of a chain is called a *class property* iff its possession by one state in a class implies its possession by all states of the same class. With these definitions, the above theorem can also be stated as follows :

In an irreducible chain all states are of the same class. The property of being recurrent (transient, persistent null, persistent non-null) is a class property. So also is the property of having a period.

Example 3(a). Consider the Markov chain with t.p.m.

$$P = \begin{array}{c|ccc} \nearrow & 0 & 1 & 2 \\ \hline 0 & 0 & 1 & 0 \\ 1 & \frac{1}{2} & 0 & \frac{1}{2} \\ 2 & 0 & 1 & 0 \end{array}.$$

The chain is irreducible as the matrix is so. We have

$$P^2 = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}, \quad P^3 = P\ ;$$

in general,

$$P^{2n} = P^2\ , \quad P^{2n+1} = P,$$

so that

$$p_{ii}^{(2n)} > 0\ , \quad p_{ii}^{(2n+1)} = 0 \quad \text{for each } i.$$

The states are periodic with period 2. It follows that $\sum_n p_{ii}^{(n)}$ diverges, so that the state i is persistent.

(Note that a periodic state must necessarily be persistent, though a persistent state may or may not be periodic.)

We find that $f_{11} = 0, f_{11}^{(2)} = 1$ so that $F_{11} = \sum_n f_{11}^{(n)} = 1$, i.e. state 1 is persistent and hence the other states 0 and 2 are also persistent.

Now

$$\mu_{11} = \sum_n n f_{11}^{(n)} = 1 \times 0 + 2 \times 1 = 2,$$

i.e. state 1 is non-null. Thus the states of the chain are periodic (each with period 2) and persistent non-null. Further,

$$p_{11}^{(2n)} \to t/\mu_{11} = 2/2 = 1 \text{ for all } n.$$

3.4 Determination of Higher Transition Probabilities

Consider a Markov chain with m states (m finite) having transition probability matrix (t.p.m.) $P = (p_{ij})$. The n-step transition probabilities $p_{ij}^{(n)}$ (i.e. the elements of P^n) can be obtained by using

$$p_{ij}^{(n)} = \sum_k p_{ik}\, p_{kj}^{(n-1)}, \quad n = 1, 2, \ldots,$$

where $$p_{jj}^{(0)} = 1, \quad p_{jk}^{(0)} = 0, \quad k \neq j,\; p_{ij}^{(1)} = p_{ij}.$$

We shall now show how some of the results of spectral theory of matrices could be used to determine $p_{ij}^{(n)}$ or P^n more expeditiously. We recall the discussions on matrices given in Sec. 1.7, Chapter 1.

Assume that the eigenvalues t_i, $i = 1, 2, \ldots, m$, of the stochastic matrix P are all distinct and different from zero. Suppose that

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})'$$

and $$\mathbf{y}_i' = (y_{i1}, y_{i2}, \ldots, y_{im})$$

are, respectively, the right and left-eigenvectors of P corresponding to t_i and that $X = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$ is the matrix formed with the eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$. Suppose that $D = (d_{ij})$ is a diagonal matrix such that

$$d_{ii} = t_i,\; d_{ij} = 0,\; j \neq i.$$

Then from the spectral theorem we have

$$P = XDX^{-1} \tag{4.1}$$

and
$$P^n = XD^nX^{-1} \tag{4.2}$$

$$= \sum_{k=1}^{m} t_k^n\, c_k\, \mathbf{x}_k \mathbf{y}_k', \text{ where } c_k = 1/(\mathbf{y}_k' \mathbf{x}_k).$$

Thus
$$p_{ij}^{(n)} = \sum_{k=1}^{m} t_k^n\, c_k\, x_{ki}\, y_{kj}, \quad n = 0, 1, 2, \ldots \tag{4.3}$$

Aperiodic Chain: Limiting Behaviour

Assume that P is primitive, i.e. the chain is aperiodic; then there will be no other eigenvalue with modulus equal to 1 except 1. Since all the eigenvalues are assumed to be distinct, we may put

$$t_1 = 1,\; |t_i| < 1,\; i \neq 1.$$

Then $x_{1i} = 1$ for all i (Theorem 1.13) and we have

$$p_{ij}^{(n)} = c_1 y_{1j} + \sum_{k=2}^{m} t_k^n\, c_k\, x_{ki}\, y_{kj}. \tag{4.4}$$

Further, in the limit, as $n \to \infty$,

$$p_{ij}^{(n)} \to c_1 y_{1j}, \quad \left(c_1 = 1/(\mathbf{y}_1' \mathbf{x}_1) = 1/\sum_{j=1}^{m} x_{1j} y_{1j}\right), \tag{4.5}$$

i.e. $P^{(n)}$ tends to a matrix with all rows equal to

$$c_1 \mathbf{y}_1' = (c_1 y_{11}, c_1 y_{12}, \ldots, c_1 y_{1m}).$$

The above shows that $\lim p_{ij}^{(n)}$ or $P^{(n)}$ exists and is independent of the initial state i, if P is primitive (see also Theorem 1.14).

To calculate this limit, when it exists, one needs determine only the left-eigenvectors $\mathbf{y}_1'$ corresponding to $t_1 = 1$.

Remark: The above method used in the study of limiting behaviour of aperiodic chains is based on spectral theory of matrices. A simpler method for deriving $\lim p_{ij}^{(n)}$ will be discussed in the next section.

Example 4(a). Consider the two-state Markov chain

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}, \; 0 < a, b < 1.$$

The eigenvalues of P are $t_1 = 1$ and $t_2 = 1 - a - b$, $| t_2 | < 1$. The right eigenvectors corresponding to $t_1 = 1$ and $t_2 = 1 - a - b$ are given, respectively, by $\mathbf{x}_1$ and $\mathbf{x}_2$, where $\mathbf{x}_1' = (1, 1)$ and $\mathbf{x}_2' = (1, -a/b)$. The left eigenvectors corresponding to t_1 and t_2 are given by $\mathbf{y}_1' = (1, a/b)$ and $\mathbf{y}_2' = (1, -1)$ respectively. We have

$$c_1 = 1/(\mathbf{y}_1'\mathbf{x}_1) = b/(a+b), \; c_2 = 1/(\mathbf{y}_2'\mathbf{x}_2) = a/(a+b),$$

$$\mathbf{x}_1\mathbf{y}_1' = \begin{pmatrix} 1 & a/b \\ 1 & a/b \end{pmatrix}, \; \mathbf{x}_2\mathbf{y}_2' = \begin{pmatrix} 1 & -1 \\ -b/a & b/a \end{pmatrix}.$$

Thus

$$P = \sum_{j=1}^{2} t_j c_j \mathbf{x}_j \mathbf{y}_j' = \frac{1}{a+b}\begin{pmatrix} b & a \\ b & a \end{pmatrix} + \frac{1-a-b}{a+b}\begin{pmatrix} a & -a \\ -b & b \end{pmatrix},$$

$$P^n = \Sigma\, t_j^n c_j \mathbf{x}_j \mathbf{y}_j' = \frac{1}{a+b}\begin{pmatrix} b & a \\ b & a \end{pmatrix}$$

$$+ \frac{(1-a-b)^n}{a+b}\begin{pmatrix} a & -a \\ -b & b \end{pmatrix}, \; n = 0, 1, 2, \ldots.$$

Further,

$$\lim_{n \to \infty} P^n = \frac{1}{a+b}\begin{pmatrix} b & a \\ b & a \end{pmatrix} = \begin{pmatrix} b/(a+b) & a/(a+b) \\ b/(a+b) & a/(a+b) \end{pmatrix}.$$

As $n \to \infty$ $\lim p_{i1} \to b/(a+b)$ and $\lim p_{i2} \to a/(a+b)$, $i = 1, 2.$ (4.6)

Example 4(b). Consider the three-state Markov chain with t.p.m.

$$P = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.4 & 0.4 \\ 0.1 & 0.5 & 0.4 \end{pmatrix}$$

The eigenvalues of P are $t_1 = 1$, $t_2 = 0.1$, $t_3 = 0.2$. The left eigenvector $\mathbf{y}_1' = (y_{11}, y_{12}, y_{13})$ corresponding to $t_1 = 1$ satisfies

$$-0.5y_{11}+0.2y_{12}+0.1y_{13}=0$$
$$0.3y_{11}-0.6y_{12}+0.5y_{13}=0$$
$$0.2y_{11}+0.4y_{12}-0.6y_{13}=0$$

We have $\mathbf{y}_1' = (0.16, 0.28, 0.24)$ and $c_1 = 1/(\sum_{j=1}^{3} y_{1j}) = 1/0.68$. Hence

$$p_{ij}^{(n)} \to c_1 y_{1j},\ (j = 1, 2, 3),$$

i.e. $$p_{i1}^{(n)} \to 0.2353,\quad p_{i2}^{(n)} \to 0.4118,\quad p_{i3}^{(n)} \to 0.3529.$$

Example 4(c). For the three-state Markov chain with t.p.m.

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

The eigenvalues are $t_1 = 1$, $t_2 = \omega$, $t_3 = \omega^2$, ω, ω^2 being the imaginary cube roots of unity. As $|\omega| = |\omega^2| = 1$, the matrix is not primitive; the chain is periodic with period 3, $(P^3 = I)$. Equation (4.5) cannot be applied in this case.

Let us now examine some of the restrictions imposed in deriving (4.1)-(4.5).

(a) The results hold even when the chain is denumerable.

(b) Suppose that 0 is an eigenvalue (say $t_m = 0$) and the other eigenvalues t_i, $i = 1, \dots, m-1$ are distinct and non-zero. Then (4.2) holds for $n = 1, 2, \dots$ with $t_m = 0$, and so (4.5) also holds.

(c) In the discussion of the spectral representation of a matrix (Sec. 1.7C) it is assumed that the eigenvalues of the matrix are all distinct. Suppose now that the characteristic equation has some multiple roots. In this case there will be some modification in the spectral representations of P and P^n given in (4.1) and (4.2).

Before the spectral representation in the general case (with multiple root) is given, the following categories of multiple roots may be noted :

(i) $t_1 = 1$ is a multiple root irrespective of whether some of the other roots t_i, $i \neq 1$ are multiple. This case which is of special importance occurs whenever the chain contains two or more closed subchains.

(ii) $t_1 = 1$ is not a multiple root but some $t_i (i \neq 1)$ are multiple roots.

Consider the general case when t_i, $i = 1, \dots, k$ are multiple roots of order $r_i (\geqslant 1)$ respectively, i.e. $\sum_{i=1}^{k} r_i = m$, the order of the chain. In this case there exists a non-singular matrix H such that P has the canonical representation,

$$P = HZH^{-1}, \tag{4.7}$$

where

$$Z = \begin{pmatrix} A_{r_1}(t_1) & & \mathbf{0} \\ & A_{r_2}(t_2) & \\ \mathbf{0} & & A_{r_k}(t_k) \end{pmatrix} \tag{4.7a}$$

and $\quad A_r(t) = (a_{i,j}(t))$ is a $(r \times r)$ matrix such that

$a_{i,j}(t) = t$ when $i = j$, $a_{i,j}(t) = 1$, when $j = i + 1$

and $\quad a_{i,j}(t) = 0$ for $j \neq i, i + 1$. Clearly, $A_1(t) = a_{1,1} = t$.

Now it can be seen that

$$(A_r(t))^n = \begin{bmatrix} t^n & \binom{n}{1} t^{n-1} & \cdots & \binom{n}{r-1} t^{n-r+1} \\ 0 & t^n & \cdots & \binom{n}{r-2} t^{n-r+2} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & t^n \end{bmatrix}, \tag{4.8}$$

where $\binom{n}{k} = 0$ for $n < k$.

We also get z^n by replacing $A_r(t)$ in the r.h.s of (4.7) by $(A_r(t))^n$. Thus

$$P^n = H Z^n H^{-1}, n = 0, 1, 2, \ldots. \tag{4.9}$$

In the general case with multiple roots we thus have (4.7) and (4.9) in place of (4.1) and (4.2) respectively.

3.5 Stability of a Markov System: Limiting Behaviour

Definition. Stationary distribution: Consider a Markov chain with transition probabilities p_{jk}. A probability distribution $\{v_j\}$ is called *stationary* (or *invariant*) for the given chain if

$$v_k = \sum_j v_j \, p_{jk}$$

such that $\quad v_j \geqslant 0, \sum_j v_j = 1.$

Again $\quad v_k = \sum_j v_j p_{jk} = \sum_j \{ \sum_i v_i p_{ij} \} \, p_{jk} = \sum_i v_i p_{ik}^{(2)}$

and, in general, $\quad v_k = \sum_i v_i p_{ik}^{(n)}, \quad n \geqslant 1.$

We now study the question—whether a Markov system, regardless of the initial state j, reaches a steady or stable state after a large number of transitions or moves. In other words, under what conditions, if any, as n tends to ∞, $p_{jk}^{(n)}$ tends to a limit v_k independent of j. When such limits exist, the system settles down and becomes stable. It then shows some long run regularity properties.

Theorem 3.6 gives a sufficient condition for the existence of v_k. It also shows how the distribution $\{v_k\}$ can be obtained, in case where this exists.

We first prove Theorem 3.5 which will be needed to prove Theorem 3.6.

Theorem 3.5. If the state j is persistent, then for every state k that can be reached from state j, $F_{jk} = 1$.

Proof: Let a_k be the probability that starting from state j, the state k is reached without previously returning to state j. The probability of never returning to state j once state k is reached is $(1 - F_{jk})$. The probability of the compound event that starting from state j, the system reaches state k (without returning to state j) and never returns to state j is $a_k (1 - F_{jk})$. If there are some other states, say $r, s, \ldots$, then we get similar terms $a_r (1 - F_{jr}), a_s (1 - F_{js}), \ldots$. Thus the probability that starting from state j, the system never returns to state j is

$$P = a_k (1 - F_{jk}) + a_s (1 - F_{js}) + a_r (1 - F_{jr}) + \ldots.$$

But since state j is persistent, $F_{jj} = 1$ and the probability of never returning to state j is $1 - F_{jj} = 0$. Thus $P = 0$. This implies that each term is zero, so that $F_{jk} = 1$. ▲

Theorem 3.6. For an irreducible ergodic chain, the limits

$$v_k = \lim_{n\to\infty} p_{jk}^{(n)} \tag{5.1}$$

exist and are independent of the initial state j. The limits v_k are such tha $v_k \geqslant 0$, $\Sigma\, v_k = 1$, i.e. the limits v_k define a probability distribution. Furthermore, the limiting probability distribution $\{v_k\}$ is identical with the stationary distribution for the given chain, so that

$$v_k = \sum_j v_j\, p_{jk}, \quad \Sigma\, v_k = 1. \tag{5.2}$$

Writing $V = (v_1, \ldots, v_k, \ldots)$, $\Sigma v_k = 1$, the relation (5.2) may also be written as

$$\begin{aligned} &V = VP \\ \text{or}\qquad &V(P - I) = 0. \end{aligned} \tag{5.3}$$

Proof: Since the states are ergodic, i.e. aperiodic, persistent non-null, $\lim\limits_{n\to\infty} p_{jk}^{(n)}$ exists and is equal to F_{jk}/μ_{kk} (Theorem 3.2). Again, by Theorem 3.5, we find that $F_{jk} = 1$, so that

$$v_k \equiv \lim_{n\to\infty} p_{jk}^{(n)} = 1/\mu_{kk} > 0,$$

and is independent of j.

We have

$$p_{jk}^{(n+m)} = \sum_i p_{ji}^{(n)}\, p_{ik}^{(m)}$$

$$\begin{aligned} v_k = \lim_{n\to\infty} p_{jk}^{(n+m)} &= \lim_{n\to\infty} \sum_i p_{ji}^{(n)}\, p_{ik}^{(m)} \\ &\geqslant \sum_i \{\lim_{n\to\infty} p_{ji}^{(n)}\}\, p_{ik}^{(m)}, \quad \text{(by Fatou's lemma} \end{aligned}$$

$$= \sum_i v_i \, p_{ik}^{(m)}, \text{ for all } m,$$

i.e.
$$v_k \geqslant \sum_i v_i \, p_{ik}^{(m)}.$$

Suppose, if possible, that $v_k > \sum_i v_i \, p_{ik}^{(m)}$; then summing over all k, we get

$$\sum_k v_k > \sum_k \sum_i v_i \, p_{ik}^{(m)} = \sum_i v_i,$$

which is impossible. Hence the sign of equality holds, i.e.

$$v_k = \sum_i v_i \, p_{ik}^{(m)}$$

for all $m \geqslant 1$. In particular, when $m = 1$

$$v_k = \sum_i v_i \, p_{ik}.$$

For large m, we have

$$v_k = \sum_i v_i v_k = (\sum_i v_i) \, v_k.$$

Hence $\Sigma \, v_i = 1$. Besides showing that $\{v_k\}$ is a probability distribution, the normalising condition $(\Sigma v_i = 1)$ ensures that the distribution is unique. ▲

We state, without proof, the *converse* which also holds.

If a chain is irreducible and aperiodic and if there exists a unique stationary distribution $\{v_k\}$ for the chain, then the chain is ergodic and $v_k = 1/\mu_{kk}$.

Thus ergodicity is a necessary and sufficient condition for the existence of $\{v_k\}$ satisfying (5.2) in case of an irreducible and aperiodic chain.

Note : The matrix $(P-I)$ in (5.3) is such that its off-diagonal elements are non-negative and the diagonal elements are strictly negative. From these properties and the irreducibility condition it follows that the matrix $(P-I)$ is of rank $(m-1)$, where m is the number of states of the chain. If the number of states is finite, then $\{v_k\}$ can be obtained from the homogeneous equations $V(P-I)=0$ (which are finite in number), and the normalising condition $\Sigma \, v_k = 1$. Some examples are considered below.

Example 5(a). Consider the two-state Markov chain of Example 4(a). The chain is irreducible and consists of two persistent non-null states. The equations (5.2) giving v_1, v_2 become

$$av_1 - bv_2 = 0 \text{ and } v_1 + v_2 = 1,$$

whence we get $v_1 = b/(a+b)$ and $v_2 = a/(a+b)$. These were directly obtained as limits in Example 4(a).

Example 5(b). Consider the three-state Markov chain having t.p.m.

$$\begin{pmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

The chain is irreducible and the states are persistent non-null. The equations (5.2) giving v_k $(k = 1, 2, 3)$ become

$$0 = -v_1 + \tfrac{1}{2} v_2 + \tfrac{1}{2} v_3$$
$$0 = \tfrac{2}{3} v_1 - v_2 + \tfrac{1}{2} v_3$$
$$1 = v_1 + v_2 + v_3.$$

We get

$$v_1 = 1/3,\ v_2 = 10/27,\ v_3 = 8/27.$$

Example 5(c). One-dimensional random walk with two reflecting barriers. Let the states be $1, 2, \ldots, k$. The particle moves from state i to state $(i+1)$ with probability p and to state $(i-1)$ with probability $q = 1-p$ for $i = 2, \ldots, k-1$. From states 1 and k the particle is reflected, i.e. from state 1 it moves to state 2 with probability p and remains at state 1 with probability q and similarly from state k it moves to state $(k-1)$ with probability q and remains at state k with probability p.

The transition probabilities are :

$$\begin{aligned} p_{i,j} &= p, \quad j = i+1 \\ &= q, \quad j = i-1 \quad \text{for} \quad 1 < i < k \\ p_{ii} &= 0, \quad i \neq 1, k \\ p_{1,2} &= p, \quad p_{1,1} = q \\ p_{k,k} &= p, \quad p_{k,k-1} = q. \end{aligned}$$

The chain is irreducible and the states are persistent non-null. The limiting probabilities v_k are solutions of the equations (5.2). The solution can be obtained recursively. From the first $(k-1)$ equations we have

$$v_2 = (p/q)v_1,\ v_3 = (p/q)^2 v_1, \ldots, v_k = (p/q)^{k-1} v_1.$$

Using $\Sigma v_j = 1$, we get

$$v_1 = (1-(p/q))/(1-(p/q)^k).$$

Hence

$$v_r = (1-(p/q))\,(p/q)^{r-1}\,(1-(p/q)^k)^{-1},\ r = 1, \ldots, k.$$

So far we have considered finite chains. If a chain is denumerable (or infinite) the number of equations given by (5.2) $(j, k = 0, 1, 2, \ldots)$ will be infinite involving an infinite number of unknowns. Then the question of solving these equations may not arise; however, by using (5.2) it is possible to derive the limiting distribution in terms of its generating function. The following example will illustrate this point.

Example 5(d) Consider a chain with an infinite number of states $(0, 1, 2, \ldots)$ having transition probability matrix

$$\begin{pmatrix} p_0 & p_1 & p_2 & p_3 & \dots \\ p_0 & p_1 & p_2 & p_3 & \dots \\ 0 & p_0 & p_1 & p_2 & \dots \\ 0 & 0 & p_0 & p_1 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix},$$

where $\Sigma p_k = 1$.

Let

$$P(s) = \sum_k p_k s^k \text{ and } V(s) = \sum_k v_k s^k$$

be the generating functions of $\{p_k\}$ and $\{v_k\}$ respectively. The chain is irreducible and aperiodic. The states are all transient, persistent null or persistent non-null according as $P'(1) > 1, = 1$ or < 1 (see Prabhu p. 63). Thus when $P'(1) < 1$, v_k's are the unique solutions of the equations (5.2). The equations (5.2) become :

$$\begin{aligned} v_0 &= p_0 v_0 + p_0 v_1 \\ v_1 &= p_1 v_0 + p_1 v_1 + p_0 v_2 \\ v_2 &= p_2 v_0 + p_2 v_1 + p_1 v_2 + p_0 v_3 \\ &\dots \quad \dots \quad \dots \quad \dots \quad \dots \end{aligned} \tag{5.4}$$

Multiplying both sides of the $(k+1)$st equation of (5.4) by s^k ($k=0, 1, 2, \dots$) and adding over k, we get

$$\begin{aligned} V(s) &= v_0 P(s) + v_1 P(s) + v_2 s P(s) + v_3 s^2 P(s) + \dots \\ &= P(s) \{v_0 + (V(s) - v_0)/s\} \end{aligned}$$

This gives

$$V(s) = v_0 (1 - s) P(s)/(P(s) - s),$$

in terms of v_0 which can be evaluated from $\Sigma v_k = 1$. We have

$$\lim_{s \to 1} \frac{V(s)}{v_0} = \lim_{s \to 1} \frac{\{1 - s\} P(s)}{P(s) - s},$$

whence

$$v_0 = 1 - P'(1) \quad (> 0).$$

Thus

$$V(s) = \frac{\{1 - P'(1)\} (1 - s) P(s)}{P(s) - s}.$$

Graph theoretic approach

A simple graph theoretic formula for finding v_k has recently been advanced by Solberg (1975). The formula, which is similar in spirit to the well-known flow-graph formulas, is conceptually simple and is claimed to have computational advantages. Further, the formula is applicable to the continuous parameter case also (see Sec. 4.5D). This approach is considered below.

A directed graph may be used as a transition diagram of a Markov chain. In other words, the states and the transitions between states of a

Markov chain may be represented by a directed graph. Here the states of the chain correspond to the points of the graph, the transitions correspond to the arcs and the transition probabilities to the arc weights.

Definition. An *intree* T_j to a specific point j in a directed graph G, is a spanning subgraph of G in which every point but j has exactly one arc emanating from it and j has none. Thus an intree to point j is just a tree with the arcs all oriented towards the point j.

The *weight* $w(T_j)$ of an intree j is defined to be the product of the weights of all arcs appearing in T_j.

We state the theorem due to Solberg (for proof refer to his paper).

Theorem 3.7. The steady-state solution of an irreducible Markov chain (discrete or continuous parameter) with a finite number of states is given by v_k,

$$v_k = c_k / \sum_j c_j,$$

where $c_j = \sum_i \omega_i(T_j)$ is the sum over all intrees to the point j in the transition diagram of the process.

The above theorem is illustrated with the help of an example. Consider the Markov chain of Example 5(b).

The intrees to the point 1 are shown below:

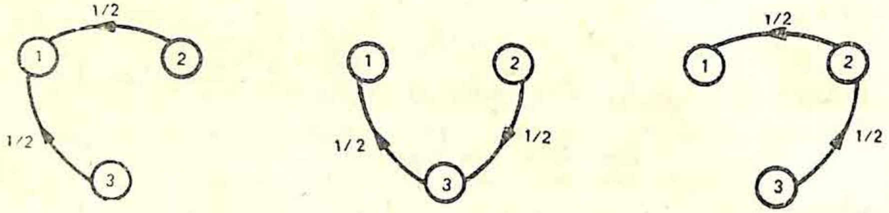


Fig. 3.1

The weights of the intrees to the point 1 are:

$$w_1(T_1) = p_{21}p_{31} = 1/4, \quad w_2(T_1) = p_{23}p_{31} = 1/4, \quad w_3(T_1) = p_{32}p_{21} = 1/4$$

so that

$$c_1 = \sum_i w_i(T_1) = 3/4.$$

The weights of the intrees to the point 2 are:

$$w_1(T_2) = p_{32}p_{12} = 1/3, \quad w_2(T_2) = p_{31}p_{12} = 1/3, \quad w_2(T_3) = p_{13}p_{32} = 1/6$$

so that

$$c_2 = \sum_i w_i(T_2) = 5/6.$$

The weights of the intrees to the point 3 are:

$$w_1(T_3) = p_{13}p_{23} = 1/6, \quad w_2(T_3) = p_{21}p_{13} = 1/6, \quad w_3(T_3) = p_{12}p_{23} = 1/3$$

so that

$$c_3 = \sum_i w_i(T_3) = 2/3.$$

Thus

$$\Sigma c_j = 3/4 + 5/6 + 2/3 = 9/4.$$

We have

$$\nu_1 = c_1/\Sigma c_j = 1/3$$

$$\nu_2 = c_2/\Sigma c_j = 10/27$$

and

$$\nu_3 = c_3/\Sigma c_j = 8/27.$$

3.6 Statistical Inference for Markov Chains

Inference problems, such as estimation and hypothesis testing, involving Markov chains have been considered by several authors (e.g. Bartlett, Whittle, Anderson and Goodman, Billingsley, and so on) not only because of their theoretical interest but also for their applications in diverse areas. Methods put forward for estimation of transition probabilities under different situations include a recent one involving linear and quadratic programming procedures to produce least squares estimates (Lee *et al.*, 1970). We shall discuss here the maximum-likelihood method of estimation of transition probabilities from individual or micro-unit data. Some tests based on these estimates will also be discussed.

Maximum-likelihood estimation: Consider a time-homogeneous Markov chain with a finite number, m, of states $(1, 2, \ldots, m)$ and having transition probability matrix $P = (p_{jk})$, $j, k = 1, 2, \ldots, m$. Suppose that the number of observed direct transitions from the state j to the state k is n_{jk}, and that the total number of observations is $(N + 1)$. Put

$$\sum_{k=1}^{m} n_{jk} = n_{j\cdot} \quad \text{and} \quad \sum_{j=1}^{m} n_{jk} = n_{\cdot k}, \quad j, k = 1, 2, \ldots, m.$$

That there is a striking similarity between a sample from a Markov chain and one from a set of independent multinomial trials has been observed, among others, by Whittle (1955) who has obtained the exact probability of the observed n_{jk} in the form:

$$f(n_{jk}) = T(n_{jk}) \frac{\prod_j (n_{j\cdot})!}{\prod_j \prod_k (n_{jk})!} \prod_j \prod_k p_{jk}^{n_{jk}}; \tag{6.1}$$

the factor $T(n_{jk})$ which involves the joint distribution of the $n_{j\cdot}$'s is independent of the p_{jk}'s.

The logarithm of the likelihood function can be put as

$$L(p_{jk}) - C + \sum_{j=1}^{m} \sum_{k=1}^{m} n_{jk} \log p_{ik}, \tag{6.2}$$

where C contains all terms independent of p_{jk}.

Since $\sum_{k} p_{jk} - 1$, (6.2) can be written as

$$L(p_{jk}) = C + \sum_{j=1}^{m} \sum_{k=1}^{m-1} n_{jk} \log p_{jk} + \sum_{j=1}^{m} n_{jm} \log \left(1 - \sum_{k=1}^{m-1} p_{jk}\right). \tag{6.3}$$

Let r be a specified value of j. The maximum-likelihood estimates $\hat{p}_{rk}$ are given by the solutions of the equations

$$\frac{\partial L(p_{rk})}{\partial p_{rk}} = 0, \qquad k = 1, 2, \ldots, (m-1).$$

These equations give

$$\frac{n_{rk}}{p_{rk}} - \frac{n_{rm}}{1 - \sum\limits_{k=1}^{m-1} p_{rk}} - 0, \quad k = 1, 2, \ldots, (m-1). \tag{6.4}$$

To fix our ideas, let us take a specified value, s, of k. Then

$$\frac{n_{rs}}{p_{rs}} = \frac{n_{rk}}{p_{rk}} = \frac{n_{rm}}{1 - \sum\limits_{k=1}^{m-1} p_{rk}}, \quad k = 1, 2, \ldots, s, \ldots (m-1).$$

Thus

$$1 - \sum_{k=1}^{m-1} p_{rk} = \frac{n_{rm}}{n_{rs}} p_{rs} \tag{6.5}$$

and

$$p_{rk} = \frac{n_{rk}}{n_{rs}} p_{rs}, \; k = 1, 2, \ldots, s, \ldots (m-1). \tag{6.6}$$

Summing (6.6) over all k and adding (6.5), we get

$$1 = \frac{\sum\limits_{k=1}^{m} n_{rk}}{n_{rs}} p_{rs}$$

and hence the estimate $\hat{p}_{rs}$ is given by

$$\hat{p}_{rs} = \frac{n_{rs}}{\sum\limits_{k=1}^{m} n_{rk}}. \tag{6.7}$$

Now r, s are two arbitrary values of j, k respectively. Hence, for $j, k = 1, 2, \ldots, (m-1)$,

$$\hat{p}_{jk} = \frac{n_{jk}}{\sum\limits_{k=1}^{m} n_{jk}} = \frac{n_{jk}}{n_{j\cdot}} \tag{6.8}$$

are the maximum-likelihood estimates of p_{jk}.

Hypothesis testing: Some of the tests developed (Anderson and Goodman) on the above estimates are given below.

(i) Suppose that one wishes to test the null hypothesis that the observed realisation comes from a Markov chain with a given transition, i.e. matrix P^0, suppose that the null hypothesis is

$$H_0 : P = P^0.$$

Then, for large N, and for $\hat{p}_{jk}$ given by (6.8), the statistic

$$\sum_{k=1}^{m} \frac{n_{j.}(\hat{p}_{jk} - p^0_{jk})^2}{p^0_{jk}}, j = 1, 2, \ldots m,$$

is distributed as χ^2 with $m - 1$ d.f. (degrees of freedom). Here p^0_{jk}'s which are equal to zero are excluded and the d.f. is reduced by the number of p^0_{jk}'s equal to 0. Alternatively, a test for all $\hat{p}_{jk}$ can be obtained by adding over all j, and the statistic

$$\sum_{j=1}^{m} \sum_{k=1}^{m} \frac{n_{j.}(\hat{p}_{jk} - p^0_{jk})^2}{p^0_{jk}} \tag{6.9}$$

has an asymptotic χ^2-distribution with $m(m - 1)$ d.f. (the number of d.f. being reduced by the number of p^0_{jk}'s equal to zero, if any, for $j, k = 1, 2, \ldots, m$).

The likelihood ratio criterion for H_0 is given by

$$\lambda = \prod_{j} \prod_{k} \left(\frac{\hat{p}_{jk}}{p^0_{jk}} \right)^{n_{jk}}$$

Thus, under the null hypothesis, the statistic

$$- 2 \log \lambda = 2 \sum_{j} \sum_{k} n_{jk} \log \frac{n_{jk}}{(n_{j.})\, p^0_{jk}} \tag{6.10}$$

has an asymptotic χ^2-distribution with $m(m - 1)$ d.f.

(ii) The maximum-likelihood estimates could also be used to test the order of a Markov chain. For testing the null hypothesis that the chain is of order zero, i.e. $H_0 : p_{jk} = p_k$ for all j, against the alternative that the chain is of order 1, the test criterion is

$$\lambda = \prod_{j} \prod_{k} \left(\frac{\hat{p}_k}{\hat{p}_{jk}} \right)^{n_{jk}},$$

where $\hat{p}_k = n_{.k}/(\sum_{j} \sum_{k} n_{jk}), \quad \hat{p}_{jk} = n_{jk}/n_{j.}$.

Under the null hypothesis, the statistic

$$- 2 \log \lambda = 2 \sum_{j} \sum_{k} n_{jk} \log \frac{n_{kj} (\sum \sum n_{jk})}{(n_{.k})(n_{j.})}$$

has an asymptotic χ^2-distribution with $(m-1)^2$ d.f. Similar test can be constructed for testing the null hypothesis that the chain is of order one against the alternative that it is of order two.

These tests do not prove that the observed chain is of order one. They provide, however, a rationale for using a Markov model. Similar likelihood ratio tests based on maximum-likelihood estimates are given (iii) for testing the stationarity (or time-homogeneity) of a Markov chain, i.e. $H_0 : p_{jk}(t) = p_{jk}$ and also (iv) for testing that several samples are from the same Markov chain of a given order.

Determination of the order of a Markov chain by MAICE

A procedure for the determination of the order of a Markov chain by Akaike's Information criterion (AIC) has been developed by Tong (1975). The AIC is defined as

AIC = (− 2) log (Maximum-likelihood) + 2 (Number of independent parameters in the model)

This statistic is introduced as a measure of deviation of the fitting model from the true structure. Given several models, the procedure envisages adoption of the model that minimises the AIC and is called Minimum AIC Estimation (MAICE). (It is argued that the MAICE procedure represents an attempt to strike a balance between overfitting which needs more parameters, and underfitting which incurs an increased residual variance.)

Denote the transition probability for a r chain by $p_{ij\ldots kl}$, $i = 1, 2, \ldots s$, s being the (finite) number of states of the chain.

Denote the ML estimates of p by $\hat{p}_{ij\ldots kl} = \dfrac{n_{ij\ldots kl}}{n_{ij\ldots k\cdot}}$,

where $n_{ij\ldots k\cdot} = \sum_{l} n_{ij\ldots kl}$. The hypothesis tested is $H_{r-1} : p_{ij\ldots kl} = p_{j\ldots kl}$, $i = 1, \ldots, s$ (that the chain is $(r-1)$-dependent against H_r : that the chain is r-dependent). The statistic constructed is

$$ {}_{r-1}A_r \equiv -2 \log \lambda_{r-1,r} = 2 \sum_{i\ldots l} (n_{ij\ldots kl}) \log \frac{(n_{ij\ldots kl})\left(\sum_{i\ldots l} (n_{ij\ldots kl})\right)}{(n_{ij\ldots k\cdot})(n_{\cdot j\ldots kl})} $$

which is a χ^2-variate with $s^{r-1}(s-1)^2$ d.f.

The hypothesis H_k that the chain is k-dependent implies the hypothesis H_r, that the chain is r-dependent for $k < r$. Denote by $\lambda_{k,r}$ the ratio of the maximum-likelihood given H_k (that the chain is of order k) to that given H_r (that the chain is of order r); then we get

$$ \lambda_{k,r} = \lambda_{k,k+1} + \ldots + \lambda_{r-1,r} $$

and so $$ {}_kA_r = -2 \log \lambda_{k,k+1} - \ldots - 2 \log \lambda_{r-1,r}, \quad k < r. $$

Assume the variables $-2 \log \lambda_{r-1,r}$ $(r = 0, 1, \ldots)$ to be asymptotically independent, given H_k; then it follows that ${}_kA_r$ has χ^2-distribution with d.f.

$$\nabla s^{r+1} - \nabla s^{k+1}\,, \quad k \geqslant 0$$

and

$$\nabla s^{r+1}\,, \qquad k = -1,$$

where

$$\nabla s^r = s^r - s^{r-1}\,, \quad r \geqslant 1.$$

Now the question of choice of an appropriate loss function arises once this identification procedure is considered as a decision procedure. The loss functions considered in classical theory of hypothesis testing are defined by the probabilities of accepting the incorrect or rejecting the correct hypothesis.

Tong proposes the choice of the loss function, based on the AIC approach as

$$R(k) = {}_kA_m - 2\,(\nabla s^{m+1} - \nabla s^{k+1}),$$

where m is the highest order model to be considered and k is the order of the fitting model. The MAICE of the best approximating order of the Markov chain is that value of k which gives the minimum of $R(k)$ over all orders considered. Note that $R(m) = 0$.

Gabriel and Neumann described the occurrences and non-occurrences of rainfall (of Tel Aviv) by a two-state Markov-chain. A dry date is denoted by state 0 and a wet date by state 1. In the following example we consider a similar model for weather data (Medhi, 1976).

Example 6(a). The following data are obtained from observations of the occurrence and non-occurrence of rainfall in Gauhati, India. A dry date is denoted by state 0 and a wet date by state 1. The period covered is from the last day of February to the last day of May (i.e. before the onset of monsoon in early June) for four consecutive years 1967-70 and N, the total number of transitions in the 4 years is $4 \times 92 = 368$. The transitions are as noted below.

↗	0	1	Total
0	175	49	224
1	48	96	144
Total	223	145	368

The estimated transition matrix $P = (p_{ij})$ is given by

$$\begin{array}{cc} & \begin{array}{cc} 0 & \quad 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 25/32 & 7/32 \\ 1/3 & 2/3 \end{pmatrix} \end{array}.$$

We have

$$P^5 = \begin{pmatrix} 0.61091 & 0.38909 \\ 0.59289 & 0.40711 \end{pmatrix},$$

$$P^{10} = \begin{pmatrix} 0.60390 & 0.39610 \\ 0.60358 & 0.39642 \end{pmatrix}, \qquad P^{20} = \begin{pmatrix} 0.60377 & 0.39623 \\ 0.60377 & 0.39623 \end{pmatrix}.$$

Using (4.6), we get $\lim_{n\to\infty} P^n \to (0.60377, 0.39623)$.

We wish to test the null hypothesis that the chain is of order zero against the alternative that it is of order one.

$$H_0 : p^0_{ij} = p_j, \quad i, j = 1, 2.$$

We have

$$\begin{aligned} -2 \log \lambda &= 2 \sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij} \log \frac{(N)\, n_{ij}}{(n_i.)\,(n.j)} \\ &= 2\left\{175 \log \frac{(368)\,(175)}{(223)\,(224)} + 49 \log \frac{(368)\,(49)}{(145)\,(224)} + 48 \log \frac{(368)\,(48)}{(223)\,(144)} \right. \\ &\qquad \left. + 96 \log \frac{(368)\,(96)}{(145)\,(144)} \right\} \\ &= 53.28. \end{aligned}$$

For 1 d.f., Pr $\{\chi^2 \geqslant 53.28\}$ being very very small the null hypothesis that the chain is of order zero is rejected. Again, we wish to test the null hypothesis that the transition probability matrix is P against the alternative $P^0 = (p^0_{ij})$,

where
$$H_0 : P^0 = \begin{pmatrix} 7/15 & 8/15 \\ 1/5 & 4/5 \end{pmatrix}$$

is the transition probability matrix of dry and wet days during the monsoon months (June-August).

We have

$$\sum_{i=1}^{2} \sum_{j=1}^{2} \frac{n_i.(p_{ij} - p^0{}_{ij})^2}{p^0{}_{ij}} = 105.07.$$

For 2 d.f., Pr $\{\chi^2 \geqslant 105.07\}$ being extremely small, the null hypothesis that $H_0 : P = P^0$ is rejected.

3.7 Markov Chains with Continuous State Space

So far we have discussed Marov chains $\{X_n, n = 0, 1, 2, \ldots\}$ with discrete state space, i.e. with $0, \pm 1, \pm 2, \ldots$ as possible values of X_n. Here we consider chains $\{X_n\}$ with continuous state space, i.e. with $(-\infty, \infty)$ as possible range of values of X_n. We shall have to use either probability distribution function (d.f.). or probability density function (p.d.f.), when this exists, in place of probability mass function.

Definition. If, for all m, and for all possible values of $\{X_m\}$ in $(-\infty, \infty)$

$$\begin{aligned} \Pr\{X_{m+1} \leqslant x \mid X_m = y, X_{m-1} = y_1, \ldots, X_0 = y_m\} \\ = \Pr\{X_{m+1} \leqslant x \mid X_m = y\} \qquad (7.1) \end{aligned}$$

then $\{X_m\}$ is said to constitute a Markov chain with continuous state space. If the conditional d.f. as given by (7.1) is independent of m, then the chain is homogeneous; (7.1) gives (one-step) transition probability d.f. More generally, the n-step transition probability d.f. is defined by

$$\begin{aligned}\Pr\{X_{m+n} \leqslant x \mid X_m = y, \ldots, X_0 = y_m\} \\ = \Pr\{X_{m+n} \leqslant x \mid X_m = y\} \\ = P_n(y; x).\end{aligned} \tag{7.2}$$

Denote
$P(y; x) = P_1(y; x) = \Pr\{X_{m+1} \leqslant x \mid X_m = y\} = \Pr\{X_{n+1} \leqslant x \mid X_n = y\}$. Let $P_n(x) = \Pr(X_n \leqslant x\}$. The transition d.f. $P_n(y; x)$ and the initial distribution $\Pr\{X_0 \leqslant x\} = P_0(x)$ can unquiely determine $P_n(x)$. The Chapman-Kolmogorov equation takes the form

$$P_{n+m}(y; x) = \int_{\infty}^{\infty} d_z P_n(y; z)\, P_m(z; x), \qquad m, n \geqslant 0, \tag{7.3}$$

which corresponds to $p_{jk}^{(n+m)} = \sum_s p_{js}^{(n)}\, p_{sk}^{(m)}$ for Markov chains with discrete state space.

For $m = 1$, we can write (7.3) as

$$\begin{aligned}P_{n+1}(y; x) &= \int d_z\, P_n(y; z)\, P_1(z; x) \\ &= \int d_z\, P_n(y; z)\, P(z; x).\end{aligned} \tag{7.4}$$

Suppose that as $n \to \infty$, $P_n(y; x)$ tends to a limit $P(x)$ independent of the initial value. Then the limiting distribution $P(x)$ satisfies the integral equation

$$\begin{aligned}P(x) &= \int d_z\, P(z)\, P(z; x) \\ &= \int P(z; x)\, dP(z).\end{aligned} \tag{7.5}$$

The relation (7.5) corresponds to (5.2) for chains with discrete state space In the above relations distribution functions can be replaced by p.d.f.'s when these exist. It is to be noted that the theory of Markov chains with continuous state space is not so fully developed as that of chains with discrete state space. Special methods are however developed to study particular types of chains such as random walks. We consider here two examples.

Example 7 (a). Random walk: Suppose that the steps of the unrestricted random walk are i.i.d. continuous random variables $Z_1, Z_2, \ldots$ with a p.d.f. $f(z)$, $-\infty < z < \infty$. We have $X_n = Z_1 + \ldots + Z_n$, so that $X_{n+1} = X_n + Z_{n+1}$. Clearly, $\{X_n\}$ is a Markov chain with a continuous state space. The (unit-step) transition d.f. can be written as

$$P(y; x) = \Pr\{X_{n+1} \leqslant x \mid X_n = y\}$$
$$= \Pr\{X_n + Z_{n+1} \leqslant x \mid X_n \leqslant y\}$$

whence we get

$$f(x) = \int f_n(z) f(x-z)\, dz,$$

where $f_n(z)$ is the n-fold convolution of $f(z)$. Here $f(x-z)$ which is the p.d.f. of $X_{n+1} = X_n + Z_{n+1}$ given that $X_n = z$, is the transition p.d.f.

Example 7 (b). Queueing process*: Suppose that customers arrive at a facility at instants t_0 ($= 0$), $t_1, t_2, \ldots$, i.e. $a_n = t_{n+1} - t_n$ is the interarrival service time between the nth and $(n+1)$ st customers. Suppose that $b_0, b_1, b_2, \ldots$ are the durations of the successive service times. Assume that a_n, b_n are sequences of independent, positive random variables with d.f. $A(x) = \Pr\{a_n \leqslant x\}$ and $B(x) = \Pr\{b_n \leqslant x\}$ respectively. Let the random variable $W_n = W(t_n - 0)$, denote the waiting time of the nth arrival. Let $X_n = b_n - a_n$ have d.f. $F(x) = \Pr\{X_n \leqslant x\}$.

We have, for $n = 0, 1, 2, \ldots$

$$\begin{aligned} W_{n+1} &= W_n + X_n, \quad \text{if} \quad W_n + X_n > 0 \\ &= 0, \qquad\qquad \text{if} \quad W_n + X_n \leqslant 0. \end{aligned} \tag{7.6}$$

It is clear that $\{W_n\}$ constitutes a Markov chain with $(0, \infty)$ as possible range of its values. We have

$$\begin{aligned} P_0(y; x) &= 0, \quad \text{if} \quad x < y \\ &= 1, \quad \text{if} \quad x \geqslant y \end{aligned}$$

and $\qquad P_n(y; x) = \Pr\{W_{m+n} \leqslant x \mid W_n = y\}$ for $n = 1, 2, \ldots.$

In particular, the unit-step transition d.f. is

$$\begin{aligned} P(y; x) = P_1(y; x) &= \Pr\{W_{n+1} \leqslant x \mid W_n = y\} \\ &= \Pr\{W_{n+1} = 0 \mid W_n = y\} + \Pr\{0 < W_{n+1} \leqslant x \mid W_n = y\} \\ &= \Pr\{y + b_n - a_n \leqslant 0\} + \Pr\{0 < y + b_n - a_n \leqslant x\} \\ &= \Pr\{y + b_n - a_n \leqslant x\} \\ &= \Pr\{b_n - a_n \leqslant x - y\} \\ &= F(x-y). \end{aligned} \tag{7.7}$$

From (7.4), we get, for $n = 0, 1, 2, \ldots$

$$P_{n+1}(y; x) = \int_0^\infty F(x-z)\, d_z\, P_n(y; z) \tag{7.8}$$

$$= F(x-z) P_n(y; z) \Big|_{z=0}^{\infty} - \int_0^\infty dF(x-z)\, P_n(y; z)$$

---

*For a description see Chapter 10.

$$= -\int_0^\infty P_n(y; z)\, dF(x-z). \tag{7.9}$$

Suppose that $\lim_{n\to\infty} P_n(y; z) = P(z)$ exists, then we get

$$P(x) = \int_0^\infty F(x-z)\, dP(z)$$

$$= -\int_0^\infty P(z)\, dF(x-z). \tag{7.10}$$

3.8 Non-stationary or Non-homogeneous Chains

The Markov chains considered so far are time-homogeneous or stationary. In other words, we have assumed that for a discrete space Markov chain $\{X_n\}$, the transition probabilities $p_{ij} = \Pr\{X_n = j \mid X_{n-1} = i\}$ (and therefore the corresponding t.p.m. $P = (p_{ij})$) are independent of n. This assumption leads to some elegant results, as already discussed. However, the model of a homogeneous or stationary chain is not appropriate in many real situations arising, for example, in epidemiology, learning theory, etc. This points to the necessity of considering also non-stationary chains, which have so long received little attention in literature at least as much as the general case is concerned.

We consider here a matrix approach developed by Harary *et al.* (1970) for finite non homogeneous chain. This makes use of the notion of a causative matrix that, when multiplies a t.p.m. yields a subsequent one.

Suppose that the transition probabilities p_{ij} are functions of n, i.e.

$$\Pr\{X_n = j \mid X_{n-1} = i\} = p_{ij}(n).$$

Then the probabilistic situation is described by a sequence of t.p.m.'s

$$P_1, P_2, \ldots, P_n, \ldots,$$

In the case of homogeneous chains, the situation is described by a single t.p.m. P, i.e. $P_n = P$ for all n A sequence of causative matrices $C_1, C_2, \ldots, C_n, \ldots$ are introduced to describe the change in the transition probability matrices. These causative matrices will be defined by

$$P_1C_1 = P_2,\ P_2C_2 = P_3,\ \ldots,\ P_nC_n = P_{n+1} \ldots \tag{8.1}$$

From (8.1) we get $C_n = P_n^{-1} P_{n+1}$, provided the matrix P_n is nonsingular. The matrices C_n multiply P_n from the right to yield P_{n+1}. Likewise one may consider causative matrices which involve multiplication from the left.

The homogeneous chain can be considered as a special case obtained by taking

$$C_n = I \quad \text{for all } n.$$

We shall assume that all the causative matrices C_n are equal, i.e. $C_n = C$ for all n. We then have

$$P_{n+1} = P_1 C^n \tag{8.2}$$

and, in general,

$$P_{n+m} = P_m C^n, \quad n = 0, 1, 2, \ldots$$

Further, if the matrices are non-singular, then

$$P_{n+1} = P_n P_{n-1}^{-1} P_n$$

and $$P_n = (P_2 P_1^{-1})^{n-2} P_2 = P_2 (P_1^{-1} P_2)^{n-2}, \quad n = 2, 3, \ldots.$$

Denote the t.p.m from period u to v by $P(u, v)$ and $P_1 = P(0, 1)$ by Q. Then

$$\begin{aligned} P(n, n+1) &= P_{n+1} \\ &= Q C^n, \quad n = 0, 1, 2, \ldots, \end{aligned}$$

and $$\begin{aligned} P(0, n+1) &= P(0, 1)\, P(1, 2) \ldots P(n, n+1) \\ &= Q.\, QC \ldots QC^n \\ &= \prod_{r=0}^{n} (QC^r) = T_n \quad \text{(say)}. \end{aligned} \tag{8.3}$$

A good deal of information about a chain can be had from the matrix C itself. The limiting properties of a constant chain (with $C_1 = C_2 = \ldots = C$) depend on the convergence of C^n and T_n as $n \to \infty$.

First we consider the following result:

Theorem 3.8. If Q and R are two stochastic matrices of the same order and if Q is non-singular, then the causative matrix $C = Q^{-1}R$ has unit row sums.

Proof: Let $\mathbf{e} = (1, 1, \ldots, 1)'$ be a column vector with unity as each of its elements. Then $R\mathbf{e} = \mathbf{e}$, $Q\mathbf{e} = \mathbf{e}$ and hence $Q^{-1}\mathbf{e} = \mathbf{e}$. Thus

$$C\mathbf{e} = Q^{-1}R\mathbf{e} = Q^{-1}\mathbf{e} = \mathbf{e}.$$

Note that C can have negative elements and so C may not be a stochastic matrix.

Some of the relevant questions are the existence of the limit and stochasticity of C_n and T_n. We examine them for the two-state non-stationary chains (for details, see Harary *et al*).

Two-state non-stationary chains

Let $P_1 = Q$ and $P_2 = R$ be given by

$$P_1 = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix}, \quad P_2 = \begin{pmatrix} c & 1-c \\ 1-d & d \end{pmatrix}$$

and the causative matrix be

$$C = \begin{pmatrix} u & 1-u \\ u-m & 1-u+m \end{pmatrix}. \tag{8.4}$$

C is a stochastic matrix iff

$$0 \leqslant u \leqslant 1,\ 0 \leqslant u - m \leqslant 1,$$

i.e. $\max(0, m) \leqslant u \leqslant \min(1, 1 + m)$. (8.5)

Since

$$C = P_1^{-1} P_2,$$

we have

$$u = (bc - (1-a)(1-d))/(a+b-1)$$

$$= (a + d - ad + bc - 1)/(a + b - 1),$$

and

$$m = (c + d - 1)/(a + b - 1).$$

The characteristic roots of C are 1 and m, so that the characteristic roots of C^n are 1 and m^n. Since $t_r(C^n)$ = sum of the characteristic roots, and C^n has unit row sums, we may write

$$C^n = \begin{pmatrix} u_n & 1 - u_n \\ u_n - m^n & 1 - u_n + m^n \end{pmatrix},\ n = 1, 2, \ldots,$$

taking $u_1 = u$. Since $C^{n+1} = C \cdot C^n = C^n . C$, we have

$$u_{n+1} = u_n - m^n + um^n = u - m + m\,u_n,\ n = 1, 2, \ldots.$$

Solving this difference equation of order 1, we get u_n. For $m = 1$,

$$u_{n+1} = u_n + u - 1$$

$$= u_{n-1} + 2(u - 1)$$

$$\ldots \qquad \ldots$$

$$= u_1 + n(u - 1)$$

or,

$$u_n = 1 + n(u - 1) \quad (\text{since } u_1 = u).$$

For $m \neq 1$, we get

$$u_n = \{u - m + m^n(1 - u)\}/(1 - m).$$

Thus, we get

$$\lim_{n \to \infty} u_n = \begin{cases} 1, & m = u = 1 \\ u - m)/(1 - m), & -1 < m < 1. \end{cases}$$

The limit does not exist otherwise. Thus, for $m = u = 1$.

$$\lim_{n \to \infty} C^n = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

When $m = u = 1$, the causative matrix C reduces to the identity matrix and the chain becomes homogeneous.

For $-1 < m < 1$,

$$\lim_{n \to \infty} C^n = \begin{pmatrix} (u - m)/(1 - m) & (1 - u)/(1 - m) \\ (u - m)/(1 - m) & (1 - u)/(1 - m) \end{pmatrix}$$

$$= ((u-m)/(1-m),\ (1-u)/(1-m)). \tag{8.6}$$

The limiting matrix is stochastic iff

$$0 \leqslant (u-m)/(1-m) \leqslant 1$$

or

$$m \leqslant u \leqslant 1; \tag{8.7}$$

when $0 \leqslant m < 1$, then (8.5) and (8.7) are equivalent, i.e. $\lim C^n$ is stochastic iff C is stochastic.

More important are the existence and stochasticity of

$$\lim_{n\to\infty} T_n = \lim_{n\to\infty} \prod_{r=0}^{n} (QC^r)$$

These are considered in the following theorem.

Theorem 3.9. Let m be the non-unit characteristic root of the causative matrix C.

If $-1 < m < 1$, $\lim\limits_{n\to\infty} T_n$ exists and is equal to

$$((u-m)/(1-m),\ (1-u)/(1-m)).$$

If $0 \leqslant m < 1$, $\lim\limits_{n\to\infty} T_n$ is stochastic iff C is stochastic.

If $-1 < m \leqslant 0$, QC^r is stochastic for all $r = 0, 1, 2, 3, \ldots$ and so are T_n and $\lim\limits_{n\to\infty} T_n$, but then C is stochastic provided only $0 \leqslant u \leqslant 1 + m$.

For the proof which involves algebra of stochastic matrices refer to Harary *et al.*

The analysis of non-stationary chains provides interesting insight into problems otherwise intractable and also furnishes basis for useful models for many situations in real life.

The elegant result concerning the asymptotic behaviour of the n-step transition probabilities of homogeneous Markov chains does not hold good in case of non-homogeneous chains. For some recent results, on such behaviour of non-homogeneous chains, reference may be made to a paper by Cohn (1976), who presents a general treatment of the asymptotic behaviour of the transition probabilities between various groups of states. For an account of the historical development of the theory of non-homogeneous chains, see E. Seneta (1973). For recent contributions refer to Kingman, (1975), Cohn (1976), Isaacson and Madsen (1976), etc. (see also Paz (1971) for an account).

EXERCISES*

3.1 A communication system transmits the two digits 0 and 1, each of them passing through several stages. Suppose that the probability that the digit which enters remains unchanged when it leaves, is p and that it changes is $q = 1 - p$. Suppose further that X_0 is the digit which enters the first stage of the system and $X_n (n \geqslant 1)$ is the digit 0 leaving the nth stage of the system. Show that $\{X_n\}$ forms a Markov chain. Find P, P^2, P^3 and calculate

---

*P denotes the transition probability matrix of the Markov chain referred to in the exercise.

$$\Pr(X_2 = 0 \mid X_0 = 1) \quad \text{and} \quad \Pr(X_3 = 1 \mid X_0 = 0).$$

3.2 Three children (denoted by 1, 2, 3) arranged in a circle play a game of throwing a ball to one another. At each stage the child having the ball is equally likely to throw it into any one of the other two children. Suppose that X_0 denotes the child who had the ball initially and $X_n (n \geqslant 1)$ denotes the child who had the ball after n throws. Show that $\{X_n\}$ forms a Markov chain. Find P. Calculate

$$\Pr(X_2 = 1 \mid X_0 = 1), \Pr(X_2 = 2 \mid X_0 = 3), \Pr(X_2 = 3 \mid X_0 = 2),$$

and also the probability that the child who had originally the ball will have it after 2 throws.

Find P if the number of children is $m (\geqslant 3)$.

3.3 Consider a sequence $\{X_n\}$ of independent coin-tossing trials with probability p for head (H) in a trial. Denote the states of X_n by states 1, 2, 3, 4, according as the trial numbers $(n - 1)$ and n result in HH, HT, TH, TT respectively. Show that $\{X_n\}$ is a Markov chain. Find P; show that $P^m (m \geqslant 2)$ is a matrix with all the four rows having the same elements.

3.4 Suppose that a fair die is tossed. Let the states of X_n be $k (= 1, 2, \ldots, 6)$, where k is the maximum number shown in the first n tosses. Find P, P^2 and P^n. Calculate $\Pr(X_2 = 6)$.

3.5 Suppose that a ball is thrown at random to one of the 3 cells. Let $X_n (n \geqslant 1)$ be said to be in state $k (= 1, 2, 3)$ if after n throws k cells are occupied. Find P and P^n.

3.6 *Occupancy problem* (Generalisation of Exercise 5): Suppose that a ball is thrown at random to one of the r cells. Let $X_n (\geqslant 1)$ be said to be in state $k (= 1, 2, \ldots, r)$ if after n throws k cells are occupied. Find P and P^2.

3.7 Find out whether the Markov chains having for transition matrices those given in Exercise 1.22 are irreducible.

3.8 A coin is tossed, p being the probability of head in a toss. Let X_n have the two states 0, 1 according as the accumulated number of heads and tails are equal or unequal. Show that the states are transient when $p \neq \frac{1}{2}$, and persistent null when $p = \frac{1}{2}$.

3.9 A particle starting from the origin moves from position j to position $(j + 1)$ with probability a_j and returns to origin with probability $(1 - a_j)$. Suppose that the states, after n moves are, 0, 1, 2,... . Show that the state 0 is recurrent iff $\lim (a_1 \ldots a_n) \to \mathbf{0}$ as $n \to \infty$.

3.10 Suppose that for a Markov chain $p_{j,j+2} = v_j$, $p_{j,0} = 1 - v_j$ and $p_{j,k} = 0$, $k \neq j + 2, 0$. Show that the state 0(X_n is at the origin after n moves) is recurrent iff $(v_0 \ldots v_{2n}) \to 0$ and non-null iff $\sum_n (v_0 \ldots v_{2n-2})$ is finite. In particular, when $v_j = e^{-aj}$ $(a > 0)$, state 0 is recurrent non-null. (In Exercises 9 and 10, the random walk is non-homogeneous.)

3.11 Show that, for a chain with a finite number, m, of states and having a doubly stochastic matrix (p_{jk}) for its transition matrix,

$$p_{jk}^{(n)} \to v_k = 1/m \text{ for all } j, k = 1, 2, \ldots, m.$$

(A non-negative square matrix is said to be *doubly stochastic* if all the row and column sums are unity.)

A *necessary* and *sufficient* condition that all the states of an irreducible and aperiodic Markov chain possessing a finite number of states are equally likely in the limit is that the transition probability matrix is doubly stochastic.

3.12 A Markov chain with states 0, 1,... has transition probabilities

$$p_{jk} = e^{-a} \sum_{r=0}^{k} \binom{j}{r} p^r (1-p)^{j-r} a^{k-r}/(k-r)!,$$

$$j, k = 0, 1, 2, \ldots, \quad 0 < p < 1.$$

Show that the limiting distribution $\{v_k\}$ is Poisson with parameter $a/(1-p)$.

3.13 *Grouping of states of a Markov chain*: Consider that three states of the Markov chain $\{X_n\}$ described in Exercise 3 above are grouped together as follows: Y_n is said to be in state 0 if both the $(n-1)$th and nth tosses result in Heads and 1 otherwise, (i.e. state 0 for HH and state 1 for HT, TH, TT).

Show that the chain $\{Y_n\}$ is non-Markovian.

If the states of a Markov chain are grouped or lumped together, the new chain does not, in general, have the Markov property.

3.14 *Inventory model with (s, S) policy*: Suppose that the stock position of an establishment is examined at epochs of time $t_1, t_2, \ldots$ and the following policy is adopted. If at epoch t_n, the stock level is less than or equal to s, procurement is immediately done to bring it to the level S and if the level is greater than s, no replenishment is done. It is supposed that the demand a_n during (t_{n-1}, t_n) obeys the probability law

$$\Pr(a_n = k) = v_k, \quad \sum_{k=0}^{\infty} v_k = 1,$$

and that $\{a_n\}$ are i.i.d. random variables. Suppose that the states of $\{X_n\}$ consist of possible values of stock size just before examination at epoch t_n, i.e. states are $S, S-1, \ldots, 0, -1, -2, \ldots$, negative values being interpreted as unfulfilled demand for stock to be fulfilled as soon as replenishment is done. Show that $\{X_n, n \geqslant 1\}$ is a Markov chain.

3.15 *Branching process* (with one type of organism or particle) : Suppose that at the end of a lifetime an organism produces similar offspring (or a particle at the end of a transformation gives rise to similar new particles) with probability $\Pr(Z=i)=p_i$, $i=0, 1, 2, \ldots$ $\sum_i p_i=1$.

Suppose that the production of offspring does not depend on the evolution of the process and that an offspring produces new offspring, independently of others and in the same way as their parents. Let $\{X_n\}$ denote the population size at the nth generation $(n \geqslant 1)$. Assume that $X_0=1$, i.e. the process starts with one ancestor. Show that $\{X_n\}$ is a Markov chain.

Let

$$P_n(s)=\sum_{j=0}^{\infty} \Pr(X=j)\, s^j$$

be the p.g.f. of X_n and

$$P(s) \equiv P_1(s)=\sum_{j=0}^{\infty} \Pr(X_1=j)\, j^s=\sum_{j=0}^{\infty} p_j s^j$$

be the p.g.f. of X_1.

($P(s)$ is also the p.g.f. of the number of offsprings per individual.) Show that

(i) $P_{n+1}(s)=P_n(P(s))=P_{n-1}(P_2(s))=\ldots=P(P_n(s))$.

(ii) $p_{j,k}=$ coefficient of s^k in $(P(s))^j$.

(iii) $p_{j,k}^{(n)}=$ coefficient of s^k in $(P_n(s))^j$.

A detailed account is given in Chapter 9.
(See also Athreya and Ney, 1972.)

3.16 *Consumer Brand switching model* (Fourt, Whitaker) : Consider that there are N brands of a consumer product competing in a market. Define

$d_i \equiv$ brand loyalty for the ith brand (equal to the proportion of consumers repurchasing the same brand i without persuation next time),

$w_j \equiv$ purchasing pressure for the jth brand (equal to the proportion of consumers who are persuaded to purchase the jth brand next time).

Then the brand switching transition matrix is given by $A=(a_{ij})$, where

$$\begin{aligned} a_{ij} &= d_i+(1-d_i)\, w_j, \qquad && j=i, j=1, \ldots, N \\ &= (1-d_i) w_j, && j \neq i, j=1, \ldots, N \\ 0 \leqslant d_i &\leqslant 1,\ 0 \leqslant w_j \leqslant 1, && \sum_{j=1}^{N} w_j=1. \end{aligned}$$

Thus

a_{ii} gives the proportion of customers d_i, remaining with the ith brand plus the proportion of disloyal customers $(1 - d_j)$ who remain with the same brand after being persuaded by the purchasing pressure w_j.

a_{ij} $(i \neq j)$ gives the proportion of disloyal customers who switch from ith to jth brand being subjected to purchasing pressure w_j. The brand share of ith brand in period t is given by the elements of t

$$Y_t = (y_{jt}),$$

where $$Y_t = Y_{t-1} A.$$

Show that the equilibrium brand shares v_j (equilibrium probabilities v_j) are given by

$$v_j = w_j(1 - d_j) \Big/ \sum_{i=1}^{N} \{w_i/(1 - d_i)\}, \qquad j = 1, \ldots N.$$

Examine the particular cases (i) $d_i = d$: (ii) $d_i = 1$.
Whitaker (1978) develops a technique of estimating brand loyalty by using least squares method to fit the above Markov model to aggregate data.

REFERENCES

T. W. Anderson and L. A. Goodman, "Statistical inference about Markov chains." *Ann. Math. Stat.*, **28**, 89-110 (1957).

K. B. Athreya and P. Ney, *Branching Processes*, Springer-Verlag, Berlin, (1972).

A.T. Bharucha-Reid, *Elements of the Theory of Markov Processes and their Applications*, McGraw-Hill, New York (1960).

P. Billingsley, "Statistical methods in Markov chains." *Ann. Math. Stat.*, **32**, 12-40 (1961).

———, *Statistical Inference for Markov Processes*, University of Chicago Press, Chicago (1961).

K. L. Chung, *Markov Chains with Stationary Transition Probabilities*, 2 ed., Springer-Verlag, New York (1967).

H. Cohn, "Finite non-homogeneous Markov chains : Asymptotic behaviour." *Adv. Appl. Prob.*, **8**, 502-516 (1976).

L. Collin, "Estimating Markov transition probabilities from micro-unit data," *Appl Stat.*, **23**, 355-71 (1974).

D. R. Cox and H. D. Miller, *The Theory of Stochastic Processes*, Methuen, London, (1965).

W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. I, 3 ed., Wiley, New York, (1968).

F. Harary, B. Lipstein and G. P. H. Styan, "A Matrix approach to nonstationary chains," *Opns. Res.*, **18**, 1168-1181 (1970).

D.L. Isaacson and R.W. Madsen, *Markov Chains—Theory and Applications*, Wiley, New York, (1976).

S. Karlin, *A First Course in Stochastic Processes*, Academic Press, New York (1966).

J.G. Kemeny and J.L. Snell, *Finite Markov Chains*, Van Nonstrand, New Jersey, (1960).

J.F.C. Kingman, "Geometrical aspects of the theory of non-homogeneous Markov chains," *Math. Proc. Camb. Phil. Soc.*, 77, 171–185 (1975).

T.C. Lee, G.G. Judge and A. Zellner, *Estimating the Parameters of the Markov Probability Model for Aggregate Time Series Data*, North Holland, Amsterdam, (1970).

J. Medhi, "A Markov chain model for the occurrences of dry and wet days," *Ind. J. Met. Hydrol. Geophys.*, **27**, 431–435 (1976).

E. Parzen, *Stochastic Processes*, Holden-Day, San Francisco, (1962).

A. Paz, *Introduction to Probabilistic Automata*, Academic Press, New York (1971).

N.U. Prabhu, *Stochastic Processes*, Macmillan, New York, (1965).

E. Seneta, "On the historical development of the theory of finite non-homogeneous Markov chains," *Proc. Camb. Phil. Soc.*, **74**, 507–513 (1973).

J.J. Solberg, "A graph theoretic formula for the steady state distribution of finite Markov processes," *Management Science*, **21**, 1040–1048 (1975).

E.J. Subelman, "On the class of Markov chains with finite convergence time," *Stoch. Proc. and Appl.*, **4**, 253–259 (1976).

H. Tong, "Determination of the order of a Markov chain by Akaike's information criterion," *J. Appl. Prob.*, **21**, 488–497 (1975).

D. Whitaker, "The derivative of a measure brand loyalty using a Markov brand switching model," *J. Op. Res. Soc.*, **29**, 959–970 (1978).

P. Whittle, " Some distributions and moment formulae for the Markov chains," *J.R.S.S* , **B 17**, 235–242 (1955).

CHAPTER 4

MARKOV PROCESSES WITH DISCRETE STATE SPACE : POISSON PROCESS AND ITS EXTENSIONS

4.1 Poisson Process

A. Introduction

Here we shall deal with some stochastic processes in continuous time and with discrete state space, which, apart from their theoretical importance, play an important role in the study of a large number of phenomena. One such process is Poisson process to which we shall confine ourselves here. To fix our ideas, consider a random event E such as (i) incoming telephone calls (at a switchboard), (ii) arrival of customers for service (at a counter), (iii) occurrence of accidents (at a certain place). Let us consider the total number $N(t)$ of occurrences of the event E in an interval of duration t, i.e. if we start from an initial epoch (or instant) $t = 0$, $N(t)$ will denote the number of occurrences upto the epoch (or instant) t (more precisely to $t + 0$). For example, if an event actually occurs at instants of time $t_1, t_2, t_3, \ldots$ then $\Lambda(t)$ jumps abruptly from 0 to 1 at $t = t_1$, from 1 to 2 at $t = t_2$ and so on; the situation can be represented graphically as in Fig. 4.1.

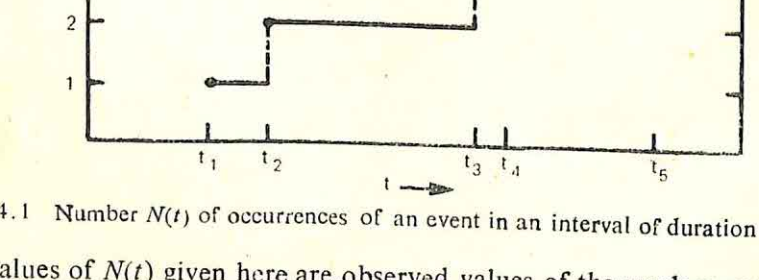


Fig. 4.1 Number $N(t)$ of occurrences of an event in an interval of duration $(0, t)$

The values of $N(t)$ given here are observed values of the random variable $N(t)$. Let $p_n(t)$ be the probability that the random variable $N(t)$ assumes the value n, i.e.

$$p_n(t) = \Pr\{N(t) = n\}. \tag{1.1}$$

This probability is a function of the time t. Since the only possible values of n are $n = 0, 1, 2, 3, \ldots,$

$$\sum_{n=0}^{\infty} p_n(t) = 1; \tag{1.2}$$

thus, $\{p_n(t)\}$ represents the probability distribution of the random variable $N(t)$ for *every value of t*. The family of random variables $\{N(t), t \geqslant 0\}$ is a stochastic process. Here the time t is continuous, the state space of $N(t)$ is discrete and integral-valued and the process is integral-valued. (such a process we also call a counting process). One of the most important integral-valued processes is Poisson process, which serves as a mathematical model for a wide range of empirical phenomena with remarkable accuracy. The justification for this is based on the concept of rare events.

We proceed to show that, under certain conditions, $N(t)$ follows Poisson distribution with mean λt (λ being a constant). In case of many empirical phenomena, these conditions are approximately true and the corresponding stochastic process $\{N(t)\}$ follows the Poisson law.

B. Postulates for Poisson Process

I. *Independence*: $N(t)$ is independent of the number of occurrences (of the event E) in an interval prior to the interval $(0, t)$, i.e. future changes in $N(t)$ are independent of the past changes.

II. *Homogeneity in time*: $p_n(t)$ depends only on the length t of the interval and is independent of where this interval is situated, i.e. $p_n(t)$ gives the probability of the number of occurrences (of E) in the interval $(t_1, t + t_1)$ (which is of length t) for every t_1.

III. *Regularity*: In an interval of infinitesimal length h, the probability of exactly one occurrence is $\lambda h + o(h)$ and that of more than one occurrence is of $o(h)$. (Occurrences are regular or orderly.)

$o(h)$ is used as a symbol to denote a function of h which tends to 0 more rapidly than h, i.e.

$$\text{as } h \to 0, \ \frac{o(h)}{h} = 0 \tag{1.3}$$

In other words, if the interval between t and $t + h$ is of short duration h, then

$$p_1(h) = \lambda h + o(h) \tag{1.4}$$

$$\sum_{k=2}^{\infty} p_k(h) = o(h) \tag{1.5}$$

Since

$$\sum_{n=0}^{\infty} p_n(h) = 1,$$

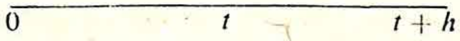
it follows that

$$p_0(h) = 1 - \lambda h + o(h). \tag{1.6}$$

Theorem 4.1. Under the postulates I, II, III, $N(t)$ follows Poisson distribution with mean λt, i.e. $p_n(t)$ is given by the Poisson law:

$$p_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \qquad n = 0, 1, 2, 3, \ldots \tag{1.7}$$

Proof: Consider $p_n(t+h)$ for $n \geqslant 0$



n events by epoch $t+h$ can happen in the following mutually exclusive ways $A_1, A_2, A_3 \ldots A_{n+1}$.

For $n \geqslant 1$

A_1: n events by epoch t and no event between t and $t+h$;

we have

$$\begin{aligned}\Pr(A_1) &= \Pr\{N(t)=n\}\Pr\{N(h)=0 \mid N(t)=n\} \qquad (1.8)\\ &= p_n(t)\,p_0(h)\\ &= p_n(t)(1-\lambda h)+o(h)\end{aligned}$$

A_2: $(n-1)$ events by epoch t and 1 event between t and $t+h$; we have

$$\begin{aligned}\Pr(A_2) &= \Pr\{N(t)=n-1\}\Pr\{N(h)=1 \mid N(t)=n-1\} \qquad (1.9)\\ &= p_{n-1}(t)\,p_1(h)\\ &= p_{n-1}(t)(\lambda h)+o(h)\end{aligned}$$

For $n \geqslant 2$

A_3: $(n-2)$ events by epoch t and 2 events between t and $t+h$;

we have $\qquad \Pr(A_3) = p_{n-2}(t)\{o(h)\} = o(h)$

And so on for A_4, A_5 etc. We thus have $\Pr(A_k) = o(h)$ for $n \geqslant k \geqslant 3$. Adding the probabilities for these mutually exclusive events, we have

$$p_n(t+h) = p_n(t)(1-\lambda h) + p_{n-1}(t)(\lambda h) + o(h),\ n \geqslant 1 \tag{1.10}$$

or,
$$\frac{p_n(t+h)-p_n(t)}{h} = -\lambda p_n(t) + \lambda p_{n-1}(t) + \frac{o(h)}{h}$$

In the limit, as $h \to 0$

$$p_n'(t) = -\lambda[p_n(t) - p_{n-1}(t)],\ n \geqslant 1 \tag{1.11}$$

For $n = 0$, we get

$$p_0(t+h) = p_0(t)\,p_0(h) = p_0(t)(1-\lambda h) + o(h)$$

or,
$$\frac{p_0(t+h)-p_0(t)}{h} = -\lambda p_0(t) + \frac{o(h)}{h}$$

whence, as $h \to 0$

$$p_0'(t) = -\lambda p_0(t) \tag{1.12}$$

Initial condition: Suppose that the process starts from scratch at time 0 (or at the origin of epoch measurement), so that $N(0) = 0$, i.e.

$$p_0(0) = 1; \quad p_n(0) = 0 \quad \text{for } n \neq 0. \tag{1.13}$$

The differential-difference equations (1.11) and (1.12) together with (1.13) completely specify the system. Their solutions give the probability distribution $p_n(t)$ of $N(t)$. The solutions are given by

$$p_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \ldots \tag{1.14}$$

(see Example 1.6(b) where the equations were considered to illustrate the use of Laplace transform in solving differential-difference equations).

We indicate here two other methods of solving these equations.

(1) The method of induction: The solution of (1.12) is given by $p_0(t) = Ce^{-\lambda t}$. Since $p_0(0) = 1$, we have $C = 1$ so that $p_0(t) = e^{-\lambda t}$. Consider (1.11) for $n = 1$. Substituting the value of p_0 and solving the equation and using (1.13), we find $p_1(t) = \lambda t \,.\, e^{-\lambda t}$. Thus (1.14) is seen to hold for $n = 0$ and 1. Assuming that it holds for $(n-1)$ it can be shown likewise that it holds for n. Hence, by induction, we get (1.14) for all n.

(2) The method of generating function: Define the probability generating function

$$P(s, t) = \sum_{n=0}^{\infty} p_n(t)s^n = \sum_{n=0}^{\infty} \Pr\{N(t) = n\}s^n = E\{s^{N(t)}\}. \tag{1.15}$$

Now
$$P(s, 0) = \sum_{n=0}^{\infty} p_n(0)s^n = p_0(0) + p_1(0)s + \ldots = 1. \tag{1.16}$$

We have
$$\frac{\partial}{\partial t} P(s, t) = \sum_{n=0}^{\infty} \frac{\partial}{\partial t} p_n(t)s^n = \sum_{n=1}^{\infty} p_n'(t)s^n + p_0'(t);$$

$$\sum_{n=1}^{\infty} p_n(t)s^n = P(s, t) - P_0(t);$$

and
$$\sum_{n=1}^{\infty} p_{n-1}(t)s^n = sP(s,t). \tag{1.17}$$

Multiplying (1.11) by s^n and adding over for $n = 1, 2, 3, \ldots$ and using (1.17), we get

$$\frac{\partial}{\partial t} P(s, t) - p_0'(t) = -\lambda[\{P(s, t) - p_0(t)\} - sP(s, t)]$$

or
$$\frac{\partial}{\partial t} P(s, t) + \lambda p_0(t) = P(s, t)\{\lambda(s-1)\} + \lambda p_0(t).$$

Thus
$$\frac{\partial}{\partial t} P(s, t) = P(s, t)\{\lambda(s-1)\}. \tag{1.18}$$

Solving (1.18), we get

$$P(s, t) = Ae^{\lambda(s-1)t}. \tag{1.19}$$

Now $P(s, 0) = 1$ from (1.16), so that $A = 1$.

Hence the p.g.f. of Poisson process is given by

$$P(s, t) = e^{\lambda t (s-1)} \tag{1.20}$$

$$= e^{-\lambda t}\left\{\sum_{n=0}^{\infty} \frac{(\lambda st)^n}{n!}\right\}, \tag{1.21}$$

so that

$$p_n(t) \equiv \text{coefficient of } s^n \text{ in } P(s, t)$$

$$= \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad n \geqslant 0. \tag{1.22}$$

Corollary 1. We have

$$E\{N(t)\} = \lambda t \tag{1.23}$$

and

$$\text{var}\,\{N(t)\} = \lambda t. \tag{1.24}$$

The mean number of occurrences in an interval of length t is λt, so that the mean number of occurrences per unit time ($t = 1$), i e. in an interval of unit length is λ. The mean rate λ per unit time is known as the *parameter* of the Poisson process.

The mean and the variance of $N(t)$ are dependent on t; in fact, its distribution is functionally dependent on t. As such the process $\{N(t)\}$ is not stationary—it is evolutionary (Sec. 2.3).

Corollary 2. If E occurred r times upto the initial instant 0 from which t is measured, then the initial condition will be

$$p_r(0) = 1, \quad p_n(0) = 0, \quad n \neq r. \tag{1.25}$$

Then $p_n(t) = \Pr$ {Number $N(t)$ of occurrences by epoch t is $n - r$, $n \geqslant r$}

$$= \frac{e^{-\lambda t}(\lambda t)^{n-r}}{(n-r)!}, \quad n \geqslant r \tag{1.26}$$

$$= 0 \qquad n < r.$$

Remarks:

Postulate I implies that the Poisson process is Markovian; postulate II that the Poisson process is time homogeneous; postulate III that in an infinitesimal interval of length h, the probability of exactly one occurrence is approximately proportional to the length h of that interval and that of the simultaneous occurrence of two (or more) events is extremely rare. In (1.8) we have put

$$\Pr\{N(h) = 0 \mid N(t) = n\} = \Pr\{N(h) = 0\} = p_0(h)$$

(here the two intervals, one of length h and the other of t, are non-overlapping). In general,

$$\Pr\{N(h) = k \mid N(t) = n\} = \Pr\{N(h) = k\} = p_k(h),\ k \geqslant 0.$$

Poisson process is a Markov process such that the conditional probabilities are constant (independent of t) and are given by (1.4)-(1.6). We shall consider time-dependent Poisson process in Sec. 4.3D. It can be easily verified that the Poisson process satisfies the Chapman-Kolmogorov equation (5.11) (see Sec. 4.5 and also Example 5(a)).

Example 1(a). Suppose that customers arrive at a Bank according to a Poisson process with a mean rate of a per minute. Then the number of customers $N(t)$ arriving in an interval of duration t minutes follows a Poisson distribution with mean at. If the rate of arrival is 3 per minute, then in an interval of 2 minutes, the probability that the number of customers arriving is

(i) exactly 4 is

$$\frac{e^{-6}(6)^4}{4!} = 0.133,$$

(ii) greater than 4 is

$$\sum_{k=5}^{\infty} \frac{e^{-6}(6)^k}{k!} = 0.714,$$

(iii) less than 4 is

$$\sum_{k=0}^{3} \frac{e^{-6}(6)^k}{k!} = 0.152,$$

(using tables of Poisson distribution).

Example 1(b). A machine goes out of order whenever a component (part fails. The failure of this part is in accordance with a Poisson process with mean rate of 1 per week.

Then the probability that two weeks have elapsed since the last failure is $e^{-2} = 0.135$, being the probability that in time $t = 2$ weeks, the number of occurrences (or failures) is 0.

Suppose that there are 5 spare parts of the component in an inventory and that the next supply is not due in 10 weeks. The probability that the machine will not be out of order in the next 10 weeks is given by

$$\sum_{k=0}^{5} \frac{e^{-10}(10)^k}{k!} = 0.068,$$

being the probability that the number of failures in $t = 10$ weeks will be less than or equal to 5.

Example 1(c). Estimation of the parameter of Poisson process. For Poisson process $N(t)$, as $t \to \infty$

$$\Pr\left\{\left|\frac{N(t)}{t}-\lambda\right|\geqslant\varepsilon\right\}\to 0, \tag{1.27}$$

where $\varepsilon > 0$ is a preassigned number.

This can be proved by applying Tshebyshev's lemma

$$\Pr\{|X-E(X)|\geqslant a\}\leqslant\frac{\operatorname{var}(X)}{a^2}, \text{ for } a>0$$

From the above, we have, for $X = N(t)$

$$\Pr\{|N(t)-\lambda t|\geqslant a\}\leqslant\frac{\lambda t}{a^2}$$

or

$$\Pr\left\{\left|\frac{N(t)}{t}-\lambda\right|\geqslant\frac{a}{t}\right\}\leqslant\frac{\lambda t}{a^2}$$

or

$$\Pr\left\{\left|\frac{N(t)}{t}-\lambda\right|\geqslant\varepsilon\right\}\leqslant\frac{\lambda}{t\varepsilon^2}$$

Hence

$$\Pr\left\{\left|\frac{N(t)}{t}-\lambda\right|\geqslant\varepsilon\right\}\to 0 \text{ as } t\to\infty.$$

This implies that for large t, the observation $N(t)/t$ may be used as a reasonable estimate of the mean rate λ of the process $\{N(t)\}$.

C. Properties of Poisson Process

1. *Additive property:* Sum of two independent Poisson processes is a Poisson process.

Let $N_1(t)$ and $N_2(t)$ be two Poisson processes with parameters λ_1, λ_2 respectively and let

$$N(t)=N_1(t)+N_2(t).$$

The p.g.f. of $N_i(t)$ $(i=1, 2)$ is

$$E\{s^{N_i(t)}\}=e^{\lambda_i(s-1)t}.$$

The p.g.f. of $N(t)$ is

$$E\{s^{N(t)}\}=E\{s^{N_1(t)+N_2(t)}\}.$$

and because of independence of $N_1(t)$ and $N_2(t)$, we have

$$\begin{aligned}E\{s^{N(t)}\}&=E\{s^{N_1(t)}\}\,E\{s^{N_2(t)}\}\\&=\{e^{\lambda_1(s-1)t}\}\,\{e^{\lambda_2(s-1)t}\}\\&=e^{(\lambda_1+\lambda_2)(s-1)t}\end{aligned}$$

Thus $N(t)$ is a Poisson process with parameter $\lambda_1+\lambda_2$.

The result can also be proved as follows :

$$\begin{aligned}\Pr\{N(t)=n\}&=\sum_{r=0}^{n}\Pr\{N_1(t)=r\}\Pr\{N_2(t)=n-r\}\\&=\sum_{r=0}^{n}\frac{e^{-\lambda_1 t}(\lambda_1 t)^r}{r!}\cdot\frac{e^{-\lambda_2 t}(\lambda_2 t)^{n-r}}{(n-r)!}\end{aligned}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)t}\{(\lambda_1+\lambda_2)t\}^n}{n!},\ n \geqslant 0. \tag{1.28}$$

Hence $N(t)$ is a Poisson process with parameter $(\lambda_1 + \lambda_2)$.

2. *Difference of two independent Poisson processes:* The probability distribution of $N(t) = N_1(t) - N_2(t)$ is given by,

$$\Pr\{N(t) = n\} = e^{-(\lambda_1+\lambda_2)t}\left(\frac{\lambda_1}{\lambda_2}\right)^{n/2} I_{|n|}\,(2t\sqrt{\lambda_1\lambda_2}),\ n = 0, \pm 1, \pm 2 \ldots, \tag{1.29}$$

where
$$I_n(x) = \sum_{r=0}^{\infty} \frac{(x/2)^{2r+n}}{r!\,\Gamma(r+n+1)} \tag{1.30}$$

is the modified Bessel function of order n $(\geqslant -1)$.

Proof: (i) The p.g.f. of $N(t)$ is

$$\begin{aligned} E\{s^{N(t)}\} &= E\{s^{N_1(t)-N_2(t)}\} \\ &= E\{s^{N_1(t)}\}\, E\{s^{-N_2(t)}\}, \end{aligned}$$

because of the independence of $N_1(t)$ and $N_2(t)$. Thus

$$\begin{aligned} E\{s^{N(t)}\} &= E\{s^{N_1(t)}\}\ E\{(1/s)^{N_2(t)}\} \\ &= \exp\{\lambda_1 t(s-1)\}\ \exp\{\lambda_2 t(s^{-1}-1)\} \\ &= \exp\{-(\lambda_1+\lambda_2)t\}\exp\{\lambda_1 ts + \lambda_2 t/s\}; \end{aligned} \tag{1.31}$$

$\Pr\{N(t) = n\}$ is given by the coefficient of s^n in the expansion of the right hand side of (1.31) as a series in positive and negative powers of s.

(ii) $\Pr\{N(t) = n\}$ can also be obtained directly as follows:

$$\begin{aligned} \Pr\{N(t) = n\} &= \sum_{r=0}^{\infty} \Pr\{N_1(t) = n + r\}\Pr\{N_2(t) = r\} \\ &= \sum_{r=0}^{\infty} \frac{e^{-\lambda_1 t}(\lambda_1 t)^{n+r}}{(n+r)!}\ \frac{e^{-\lambda_2 t}(\lambda_2 t)^r}{r!} \\ &= e^{-(\lambda_1+\lambda_2)t}\left(\frac{\lambda_1}{\lambda_2}\right)^{n/2} \sum_{r=0}^{\infty} \frac{(t\sqrt{\lambda_1\lambda_2})^{2r+n}}{r!(r+n)!} \\ &= e^{-(\lambda_1+\lambda_2)t}\left(\frac{\lambda_1}{\lambda_2}\right)^{n/2} I_{|n|}\,(2t\sqrt{\lambda_1\lambda_2}). \end{aligned}$$

It may be noted that :

(1) the difference of two independent Poisson processes is *not* a Poisson process;

(2) $I_{-n}(t) = I_n(t) = I_{|n|}(t),\ n = 1, 2, 3, \ldots$

(3) the first two moments of $N(t)$ are given by

$$E\{N(t)\} = (\lambda_1 - \lambda_2)t \quad E\{N^2(t)\} = (\lambda_1 + \lambda_2)t + (\lambda_1 - \lambda_2)^2 t^2.$$

Example 1(d). If passengers arrive (singly) at a taxi stand in accordance with a Poisson process with parameter λ_1 and taxis arrive in accordance with a Poisson process with parameter λ_2 then $N(t) = N_1(t) - N_2(t)$ gives the excess of passengers over taxis in an interval t. The distribution of $N(t)$, i.e., $\Pr\{N(t) = n\}$, $n = 0, \pm 1, \pm 2, \ldots$ is given by (1.29). The mean of $N(t)$ is $(\lambda_1 - \lambda_2)t$, which is $>, =$ or < 0 according as $\lambda_1 >, =$ or $< \lambda_2$, and var $\{N(t)\} = (\lambda_1 + \lambda_2)t$.

3. *Decomposition of a Poisson process:* A *random* selection from a Poisson process yields a Poisson process. Suppose that $N(t)$, the number of occurrences of an event E in an interval of length t is a Poisson process with parameter λ. Suppose also that each occurrence of E has a constant probability p of being recorded, and that the recording of an occurrence is independent of that of other occurrences and also of $N(t)$.

If $M(t)$ is the number of occurrences recorded in an interval of length t, then $M(t)$ is also a Poisson process with parameter λp.

Proof: The event $\{M(t) = n\}$ can happen in the following mutually exclusive ways :

A_r : E occurs $(n+r)$ times by epoch t and exactly n out of $(n+r)$ occurrences are recorded, probability of each occurrence recorded being p, $(r = 0, 1, 2, \ldots)$.

We have

$\Pr(A_r) = \Pr\{E$ occurs $(n+r)$ times by epoch $t\}.\Pr\{n$ occurrences are recorded given that the number of occurrences is $n+r\}$

$$= \frac{e^{-\lambda t}\,(\lambda t)^{n+r}}{(n+r)\,!}\binom{n+r}{n}p^n q^r.$$

Hence

$$\begin{aligned}
\Pr\{M(t) = n\} &= \sum_{r=0}^{\infty} \Pr(A_r)\\
&= \sum_{r=0}^{\infty} \frac{e^{-\lambda t}\,(\lambda t)^{n+r}}{(n+r)\,!}\binom{n+r}{n}p^n q^r\\
&= e^{-\lambda t}\sum_{r=0}^{\infty}\frac{(\lambda pt)^n\,(\lambda qt)^r}{n!\,r!}\\
&= e^{-\lambda t}\frac{(\lambda pt)^n}{n!}\sum_{r=0}^{\infty}\frac{(\lambda qt)^r}{r!}\\
&= e^{-\lambda t}\frac{(\lambda pt)^n}{n!}e^{\lambda qt} = \frac{e^{-\lambda pt}\,(\lambda pt)^n}{n!}. \quad \blacktriangle \qquad (1.32)
\end{aligned}$$

We can interpret the above as follows :

For a Poisson process $N(t)$, the probability of an occurrence in an infinitesimal interval h is (approximately) proportional to the length of the interval h, the constant of proportionality being λ.

Now for $M(t)$, the probability of a recording in the interval h is proportional to the length h, the constant of proportionality being λp. Thus $\{M(t)\}$ is a Poisson process with parameter $\lambda\, p$.

4. *(Continuation of property 3) :* The number $M_1(t)$ of occurren-

ces not recorded is also a Poisson process with parameter $\lambda q = \lambda(1-p)$ and $M(t)$ and $M_1(t)$ are independent.

Thus by random selection a Poisson process $\{N(t)\}$ of parameter λ is decomposed into two independent Poisson processes $\{M(t)\}$ and $\{M_1(t)\}$ with parameters λp and $\lambda(1-p)$ respectively.

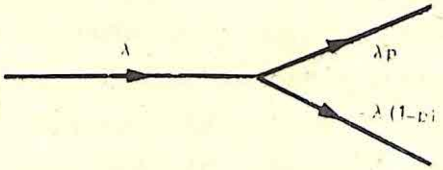


Fig. 4 2

As an example, suppose that the births (in a certain town hospital) occur in accordance with a Poisson process with parameter λ. If the probability that an individual born is male is p, then the male births form a Poisson process with parameter λp and the female births form an independent Poisson process with parameter $\lambda(1-p)$.

More generally, a Poisson process $N(t)$ with parameter λ may be decomposed into r streams of Poisson processes. If $p_1, p_2, \ldots, p_r$, are the probability of the process decomposing into r independent streams such that $p_1 + \ldots + p_r = 1$ then the Poisson process is decomposed into r independent Poisson processes with parameters $\lambda p_1, \lambda p_2, \ldots, \lambda p_r$.

5. *Poisson process and binomial distribution :* If $N(t)$ is a Poisson process and $s < t$, then

$$\Pr\{N(s) = k \mid N(t) = n\} = \binom{n}{k} (s/t)^k (1-(s/t))^{n-k}.$$

Proof:

$$\begin{aligned}\Pr\{N(s) = k \mid N(t) = n\} &= \frac{\Pr\{N(s) = k \text{ and } N(t) = n\}}{\Pr\{N(t) = n\}}\\ &= \frac{\Pr\{N(s) = k \text{ and } N(t-s) = n-k\}}{\Pr\{N(t) = n\}}\\ &= \frac{\Pr\{N(s) = k\}\ \Pr\{N(t-s) = n-k\}}{\Pr\{N(t) = n\}}\\ &= \frac{e^{-\lambda s}(\lambda s)^k}{k!} \cdot \frac{e^{-\lambda(t-s)}[\lambda(t-s)]^{n-k}}{(n-k)!} \Big/ \frac{e^{-\lambda t}(\lambda t)^n}{n!}\\ &= \frac{n!}{k!\,(n-k)!}\, \frac{s^k (t-s)^{n-k}}{t^n}\\ &= \binom{n}{k} (s/t)^k (1-(s/t))^{n-k}. \qquad (1.33)\end{aligned}$$

6. If $\{N(t)\}$ is a Poisson process then the (correlation) auto-correlation coefficient between $N(t)$ and $N(t+s)$ is $\{t/(t+s)\}^{1/2}$.

Proof: Let λ be the parameter of the process: then

$$E\{N(T)\} = \lambda T, \ \text{var}\{N(T)\} = \lambda T,$$

and $$E\{N^2(T)\} = \lambda T + (\lambda T)^2 \text{ for } T = t \text{ and } t + s.$$

Since $N(t)$ and $N(t+s)-N(t)$ are independent, $\{N(t)\}$ being a Poisson process,

$$\begin{aligned} E\{N(t)N(t+s)\} &= E[N(t)\{N(t+s)-N(t)+N(t)\}] \\ &= E[N(t)N(t)+N(t)\{N(t+s)--N(t)\}] \\ &= E\{N^2(t)\}+E\{N(t)\}E\{N(t+s)-N(t)\}. \end{aligned}$$

Hence $E\{N(t)N(t+s)\} = (\lambda t+\lambda^2t^2)+\lambda t.\lambda s.$

Thus the autocovariance between $N(t)$ and $N(t+s)$ is given by

$$\begin{aligned} C(t, t+s) &= E\{N(t)N(t+s)\}-E\{N(t)\}E\{N(t+s)\} \\ &= (\lambda t+\lambda^2t^2+\lambda^2ts)-\lambda t(\lambda t+\lambda s) = \lambda t. \end{aligned}$$

Hence the autocorrelation function

$$\begin{aligned} \rho(t, t+s) &= \frac{C(t, t+s)}{\{\text{var } N(t) \text{ var } N(t+s)\}^{1/2}} \\ &= \{(t/(t+s))\}^{1/2}. \end{aligned} \tag{1.34}$$

It can be shown that $\rho(t, t') = \dfrac{\min(t, t')}{(tt')^{1/2}}$. (1.35)

This is the autocorrelation function of the process.

Example 1(e). A radioactive source emits particles at a rate of 5 per minute in accordance with a Poisson process. Each particle emitted has a probability 0.6 of being recorded. The number $N(t)$ of particles recorded in an interval of length t is a Poisson process with rate $5\times 0.6 = 3$ per minute, i.e. with mean $3t$ and variance $3t$. In a 4-minute interval the probability that the number of particles recorded is 10 is equal to $e^{-12}(12)^{10}/10! = 0.104$.

Example 1(f). A person enlists subscriptions to a magazine, the number enlisted being a Poisson process with mean rate 6 per day. Subscribers may subscribe for 1 or 2 years independently of one another with respective probabilities $\frac{2}{3}$ and $\frac{1}{3}$.

The number of subscribers $N(t)$ enrolled by the person in time t days is a Poisson process with mean rate $6t$: the number $N_1(t)$ enrolled for 1 year period is a Poisson process with mean $(6\times\frac{2}{3})t = 4t$ and the number $N_2(t)$ enrolled for 2-year period is a Poisson process with mean $(6\times\frac{1}{3})t = 2t$.

If he received commission a for a 1-year subscription and b for a 2-year one, then the total commission earned in period t is given by

$$X(t) = aN_1(t)+bN_2(t).$$

We have

$$\begin{aligned} E\{X(t)\} &= aE\{N_1(t)\}+bE\{N_2(t)\} \\ &= 4at+2bt \end{aligned}$$

and

$$\text{var } X(t) = 4a^2t+2b^2t.$$

Note: The process $X(t)$ is a compound Poisson process (see Sec. 4.3B).

4.2 Poisson Process and Related Distributions

A. Interarrival Time

With a Poisson process, $\{N(t)\}$ the number of occurrences of an event E by epoch t, there is associated a random variable—the interval X between two successive occurrences of E. We proceed to show that X has a negative exponential distribution.

Theorem 4.2. The interval between two successive occurrences of a Poisson process $\{N(t)\}$ having parameter λ has a negative exponential distribution with mean $1/\lambda$.

Proof: Let X be the random variable representing the interval between two successive occurrences of $\{N(t)\}$ and let $\Pr(X \leqslant x) = F(x)$ be its distribution function.

Let us denote two successive events by E_i and E_{i+1} and suppose that E_i occurred at the instant t_i. Then

$$\begin{aligned}
\Pr\{X > x\} &= \Pr\{E_{i+1} \text{ did not occur in } (t_i, t_i + x) \text{ given that } E_i \text{ occurred} \\
&\qquad \text{at the instant } t_i\} \\
&= \Pr\{E_{i+1} \text{ did not occur in } (t_i, t_i + x) \mid N(t_i) = i\} \\
&\qquad \text{(because of the postulate of independence)} \\
&= \Pr\{\text{no occurrence takes place in an interval } (t_i, t_i + x) \text{ of} \\
&\qquad \text{length } x \mid N(t_i) = i\} \\
&= \Pr\{N(x) = 0 \mid N(t_i) = i\} = p_0(x) = e^{-\lambda x},\ x > 0. \qquad (2.1)
\end{aligned}$$

Since i is arbitrary, we have for the interval X between *any two* successive occurrences,

$$F(x) = \Pr\{X \leqslant x\} = 1 - \Pr\{X > x\} = 1 - e^{-\lambda x},\ x > 0.$$

The density function is

$$f(x) = F'(x) = \lambda e^{-\lambda x}, \quad (x > 0). \quad \blacktriangle \qquad (2.2)$$

It can further be proved that if X_i denotes the interval between E_i and E_{i+1}, $i = 1, 2, \ldots$, then $X_1, X_2 \ldots$ are also *independent*. We omit the proof which is outside the scope of this book. We state the result as follows:

Theorem 4.3. The intervals between successive occurrences (called interarrival times) of a Poisson process (with mean λt) are identically and independently distributed random variables which follow the negative exponential law with mean $1/\lambda$.

The converse also holds; this is given in Theorem 4.4 below. These two theorems give a characterisation of the Poisson process.

Theorem 4.4. If the intervals between successive occurrences of an event E are independently distributed with a common exponential distribution with mean $1/\lambda$, then the events E form a Poisson process with mean λt.

Proof: Let Z_n denote the interval between $(n-1)$th and nth occurrence of a process $\{N(t)\}$ and let the sequence $Z_1, Z_2, \ldots$ be independently and identically distributed random variables having negative exponential distribution with mean $1/\lambda$. The sum $W_n = Z_1 + \ldots + Z_n$ is the waiting time upto the nth occurrence, i.e. the time from the origin to the nth subsequent occurrence. W_n has a gamma distribution with parameters λ, n (Theorem 1.10). The p.d.f. $g(x)$ and the distribution function F_{W_n} are given respectively by

$$g(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)}, \quad x > 0$$

and
$$F_{W_n}(t) = \Pr\{W_n \leqslant t\} = \int_0^t g(x)\,dx.$$

The events $\{N(t) < n\}$ and $\{W_n = Z_1 + \ldots + Z_n > t\}$ are equivalent. Hence the distribution functions $F_{N(t)}$ and $F_{W_n}(t)$ satisfy the relation

$$\begin{aligned} F_{W_n}(t) &= \Pr\{W_n \leqslant t\} = 1 - \Pr\{W_n > t\} \\ &= 1 - \Pr\{N(t) < n\} = 1 - \Pr\{N(t) \leqslant (n-1)\} \\ &= 1 - F_{N(t)}(n-1). \end{aligned}$$

Hence the distribution function of $N(t)$ is given by

$$\begin{aligned} F_{N(t)}(n-1) &= 1 - F_{W_n}(t) \\ &= 1 - \int_0^t \frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)}\,dx \\ &= 1 - \frac{1}{\Gamma(n)} \int_0^{\lambda t} y^{n-1} e^{-y}\,dy \\ &= \frac{1}{\Gamma(n)} \int_{\lambda t}^{\infty} y^{n-1} e^{-y}\,dy \\ &= \sum_{j=0}^{n-1} \frac{e^{-\lambda t}(\lambda t)^j}{j!} \quad \text{(integrating by parts).} \end{aligned}$$

Thus the probability law of $N(t)$ is

$$\begin{aligned} p_n(t) &= \Pr\{N(t) = n\} = F_{N(t)}(n) - F_{N(t)}(n-1) \\ &= \sum_{j=0}^{n} \frac{e^{-\lambda t}(\lambda t)^j}{j!} - \sum_{j=0}^{n-1} \frac{e^{-\lambda t}(\lambda t)^j}{j!} \\ &= \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad n = 0, 1, 2, . \end{aligned}$$

Thus the process $N(t)$ is a Poisson process with mean λt. ▲

Note that Poisson process has independent exponentially distributed interarrival times and gamma distributed waiting times.

Example 2(a). Suppose that customers arrive at a counter in accordance with a Poisson process with mean rate of 2 per minute ($\lambda = 2$/minute). Then the interval between any two successive arrivals follows exponential distribution with mean $1/\lambda = \frac{1}{2}$ minute. The probability that the interval between two successive arrivals is

(i) more than 1 minute is

$$\Pr(X > 1) = e^{-\frac{1}{2}} = 0.607$$

(ii) 4 minutes or less is

$$\Pr(X \leqslant 4) = 1 - e^{-4.\frac{1}{2}} = 1 - e^{-2} = 0.865$$

(iii) between 1 and 2 minutes is

$$\Pr(1 \leqslant X \leqslant 2) = \int_1^2 \tfrac{1}{2} e^{-x/2}\, dx = 0.239.$$

Example 2(b). Suppose that customers arrive at a counter independently from two different sources. Arrivals occur in accordance with a Poisson process with mean rate of λ per hour from the first source and μ per hour from the second source. Since arrivals at the counter (from whatever source) constitute a Poisson process with mean $(\lambda + \mu)$ per hour, the interval between any two successive arrivals has a negative exponential distribution with mean $\{1/(\lambda + \mu)\}$ hours.

For example, if taxis arrive at a spot from the north at the rate of 1 per minute and from the south at the rate of 2 per minute in accordance with two independent Poisson processes, the interval between arrival of two taxis has a (negative) exponential distribution with mean $\frac{1}{3}$ minute. The probability that a lone person will have to wait more than a given time t can be found.

The Poisson type of occurrences are also called purely random events and the Poisson process is called a *process of purely random events.* The reason for this is that the occurrence is equally likely to happen anywhere in $[0, T]$ given that only one occurrence has taken place in that interval. We state this by the following theorem.

Theorem 4.5. Given that only one occurrence of a Poisson process $N(t)$ has occurred by epoch T, then the distribution of the time interval γ in $[0, T]$ in which it occurred is uniform in $[0, T]$, i.e.

$$\Pr\{t < \gamma < t + dt \mid N(T) = 1\} = dt/T, \quad 0 < t < T$$

Proof: We have

$$\Pr\{t < \gamma < t + dt\} = \lambda e^{-\lambda t}\, dt,$$

$$\Pr\{N(T) = 1\} = e^{-\lambda T}(\lambda T),$$

and

$$\Pr\{N(T) = 1 \mid \gamma = t\} = e^{-\lambda(T-t)},$$

the last one being the probability that there was no occurrence in the interval of length $(T-t)$.

Hence

$$\begin{aligned}\Pr\{t<\gamma<t+dt \mid N(T)=1\} &= \Pr\{t<\gamma<t+dt \text{ and } N(T)=1\}/\Pr\{N(T)=1\}\\ &= \Pr\{t<\gamma<t+dt\}\Pr\{N(T)=1 \mid \gamma=t\}/\Pr\{N(T)=1\}\\ &= \lambda e^{-\lambda t}dt\, e^{-\lambda(T-t)}/e^{-\lambda T}(\lambda T)=dt/T.\end{aligned}$$

It may be said that a Poisson process distributes points at random over the infinite interval $[0, \infty]$ in the same way as the uniform distribution distributes points at random over a finite interval $[a, b]$.

For a more general form of Theorem 4.5, see Exercise 4.13 (proof is similar).

B. Some Interesting Properties of Poisson Process

We consider here some interesting properties which are of great use in the study of various phenomena.

We have shown that the interval X_i $(=t_{i+1}-t_i)$ between two successive occurrences E_i, E_{i+1} $(i \geqslant 1)$ of a Poisson process with parameter λ has an exponential distribution with mean $1/\lambda$. Further, the following result holds.

I. For a Poisson process with parameter λ, the interval of time X upto the *first* occurrence also follows an exponential distribution with mean $1/\lambda$. For,

$$\Pr\{X_0 > x\} = \Pr\{N(x)=0\} = e^{-\lambda x}, \quad (x>0).$$

In other words, the relation (2.1) does not depend on i nor on t_i.

The same argument also leads to another very important property.

Suppose that the interval X is measured from an *arbitrary* instant of time $t_i+\gamma$ (γ arbitrary) in the interval (t_i, t_{i+1}) and not just the instant t_i of the occurrence of E_i, and Y is the interval upto the occurrence of E_{i+1} measured from $t_i+\gamma$, i.e.

$$Y = t_{i+1}-(t_i+\gamma);$$

Y is called *random modification* of X or *residual time* of X. It follows that:

II. If X is exponential then its random modification Y has also exponential distribution with the same mean.

We may state the result also as follows.

III. For a Poisson process with parameter λ, the interval upto the occurrence of the next event measured from any start of time (not necessarily from the instant of the previous occurrence) is independent of the elapsed time (since the previous instant of occurrence) and is a random variable having exponential distribution with mean $1/\lambda$.

The implication of the above would be clear when viewed in the light of an illustrative example (see, Karlin, p. 176) of fishing (incidentally, *poisson* is the French word for fish) as a Poisson process (named after the celebrated French mathematician Denis Poisson (1781-1840)). Suppose that the random variable $N(t)$ denotes the number of fish caught by an angler in $[0, t]$. Under certain "ideal conditions", such as (i) the number of fish available is very large, (ii) the angler stands in no better chance of catching fish than others, and (iii) the number of fish likely to nibble at one particular instant is the same as at another instant, the process $\{N(t), t \geqslant 0\}$ may be considered as a Poisson process. The interval upto the first catch, as also the interval between two successive catches has the same exponential distribution. So also is the time interval upto the next catch (from an arbitrary instant γ) which is independent of the elapsed time since the last catch to that instant γ. The long time spent since the last catch gives "no premium for waiting" so far as the next catch is concerned.

C. Geometric and Exponential Distributions

The geometric and exponential distributions are unique in the sense that these are the only distributions with memoryless or no-aging property. In fact, there is some close connection between these two distributions.

We show below how one can get the exponential distribution from the geometric distribution.

Suppose that the occurrence (or non-occurrence) of an event E may happen only at the discrete time points $\delta, 2\delta, 3\delta, \ldots$, each subinterval being of length δ, the probability of occurrence being $p = \lambda\delta$ and that of non-occurrence being $q = 1 - \lambda\delta$.

The number of failures N preceding the first success has a geometric distribution

$$\Pr\{N = k\} = q^k p, \quad k = 0, 1, 2, \ldots$$

Now the event that the first success occurs at trial number $(k+1)$ or beyond has the probability $\sum_{r=k}^{\infty} q^r p = q^k p/(1-q) = q^k$, and therefore the probability that the time T required for the first success exceeds $k\delta$ is given by

$$\Pr(T > k\,\delta) = \sum_{r=k}^{\infty} q^r p = q^k = (1 - \lambda\delta)^k.$$

Now suppose that $\delta \to 0$ as $k \to \infty$, such that $k\delta = t$ remains fixed, then as $k \to \infty$,

$$\Pr(T > t) = (1 - \lambda t/k)^k = \{(1 - \lambda t/k)^{k/\lambda t}\}^{\lambda t} \to \{e^{-1}\}^{\lambda t} = e^{-\lambda t}.$$

Thus T has an exponential distribution with mean

$$\frac{1}{\lambda} = \frac{\delta}{p}.$$

Example 2(c). Poisson process and geometric distribution. Consider two independent series of events E and F occurring in accordance with Poisson processes with mean at and bt respectively. The number N of occurrences of E between two successive occurrences of F has a geometric distribution.

Proof: The interval between two successive occurrences of F has the density $f(x) = be^{-bx}$ and therefore $\Pr(N = k)$, the probability that k occurrences of E take place during an arbitrary interval (between two successive occurrences of F) is given by

$$\Pr\{N = k\} = \int_0^\infty \frac{e^{-at}(at)^k}{k!} f(t)dt$$

$$= \int_0^\infty \frac{e^{-at}(at)^k}{k!} be^{-bt}dt$$

$$= \frac{ba^k}{k!}\int_0^\infty t^k e^{-(a+b)t}dt$$

$$= \frac{ba^k}{(a+b)^{k+1}}$$

$$= \frac{b}{a+b}\left\{\frac{a}{a+b}\right\}^k, \quad k = 0, 1, 2, 3, \ldots \quad \blacktriangle \tag{2.3}$$

Poisson Count Process: Let E and E' be two random sequences of events occurring at instants $(t_1, t_2, \ldots)$ and $(t_1', t_2', \ldots)$ respectively. The number N_n of occurrences of E' during the interval (t_{n-1}, t_n) of length $t_n - t_{n-1}$ is called the count process of E' in E. If E is Poisson then such a process is called *Poisson count process*. We have proved above that if E' is also Poisson, then the count process N_n has geometric distribution. It can also be shown that N_n $(n = 1,2,\ldots)$ are *independently* and *identically* distributed with a common geometric distribution. For more general results and for applications of Poisson count process in Queueing theory, see Kingman (1963).

Example 2(d). Suppose that E and F occur independently and in accordance with Poisson processes with parameters a and b respectively.

The time between every two occurrences of F is the sum of two independent exponential distributions and thus has the density

$$f(x) = b^2xe^{-bx}.$$

The probability that k occurrences of E take place between every second occurrence of F is given by

$$\int_0^\infty \frac{e^{-at}(at)^k}{k!} b^2te^{-bt}dt = \frac{a^k\,b^2}{k!}\int_0^\infty e^{-(a+b)t}t^{k+1}dt$$

$$= \frac{a^k\ b^2}{k!} \cdot \frac{\Gamma(k+2)}{(a+b)^{k+2}}$$

$$= (k+1)\,(b/(a+b))^2\,(a/(a+b))^k$$

$$= \binom{k+1}{1}\left(\frac{b}{a+b}\right)^2\left(\frac{a}{a+b}\right)^k, \quad k = 0, 1, 2, \ldots$$

The distribution is negative binomial. This also follows from the fact that we have here a sum of two independent geometric distributions; we have two successive intervals (of occurrences of F) in which the number of occurrences of E follows a geometric distribution (see Example 1(g), Ch. 1).

4.3 Generalisations of Poisson Process

There are several directions in which the Poisson process (which may be referred to also as the *classical* Poisson process) discussed in the previous section can be generalised. We consider some of them here.

A. Poisson Process in Higher Dimensions

We have considered so far the one-dimensional case: the occurrences take place at random instants of time t (say, $t_1, t_2, \ldots$) and thus we were concerned with distribution of points on a line. Instead, we may have the two-dimensional case.

Consider the two-dimensional case, such that for the number $N(\Delta a)$ of occurrences in an element of area Δa, we have, for infinitesimal Δa,

$$\Pr\{N(\Delta a) = 1\} = \lambda\,\Delta a + o(\Delta a).$$

$$\Pr\{N(\Delta a) = k\} = o(\Delta a), \quad k \geqslant 2$$

and

$$\Pr\{N(\Delta a) = 0\} = 1 - \lambda\,\Delta a + o(\Delta a). \tag{3.1}$$

Thus, if the number of occurrences in non-overlapping areas are mutually independent, the number $N(a)$ of occurrences in an area a will be a Poisson process with mean λa. Here in place of one-dimensional t, we consider two-dimensional a. Similarly, we can describe Poisson process in higher dimensions.

B. Poisson Cluster Process (Compound Poisson Process)

We considered in Sec. 4.1 that only one event can occur at an instant. Now let us suppose that several events can happen simultaneously at an instant, i.e. we have a cluster (of occurrences) at a point. We assume that:

(i) The number $N(t)$ of clusters in time t, i.e. the points at which a cluster occurs constitute a Poisson process with mean rate λ.

(ii) Each cluster has a random number of occurrences, i.e. the number X_i of occurrences in ith cluster is a r.v. The various numbers of occurrences in the different clusters are mutually independent and follow the same probability distribution:

$$\Pr\{X_i = k\} = p_k,\ k = 1, 2, 3, \ldots \qquad (3.2)$$
$$i = 1, 2, 3, \ldots$$

having p.g.f.

$$P(s) = \sum_{k=1}^{\infty} p_k s^k. \qquad (3.3)$$

Theorem 4.6. If $M(t)$ denote the total number of occurrences in an interval of length t under the conditions (i) and (ii) stated above, then the generating function of $M(t)$ is given by

$$G(P(s)) = \exp\,[\lambda t\,\{P(s) - 1\}]. \qquad (3.4)$$

Proof : $M(t)$ is the sum of a random number of terms, i.e.

$$M(t) = \sum_{i=1}^{N(t)} X_i,$$

where $N(t)$ is a Poisson process with mean λt.

Now $P(s)$ is the p.g.f. of X_i and $G(s)$ is the p.g.f. of $N(t)$. Thus,

$$G(s) = \exp\,\{\lambda t(s - 1)\}.$$

Hence by Theorem 1.3.,the p.g.f. of $M(t)$ is given by

$$G(P(s)) = \exp\,\{\lambda t\,(P(s) - 1)\}. \ \blacktriangle$$

Note: $M(t)$ is also called a *compound Poisson process.* It is to be noted that $M(t)$ is not necessarily Poisson. Poisson cluster processes (compound Poisson processes) arise in bulk queues, where customers arrive or are served in groups. In fact, a large variety of practical problems can be reduced to compound Poisson process (Feller II, p. 179).

An important application arises in '*Collective risk theory*'. Suppose that claims against a company occur in accordance with a Poisson process with mean λt, and that individual claims X_i are i.i.d. with distribution $\{p_k\}$, then $M(t)$ represents the total claims at epoch t. If A represents initial reserve and c the rate of increase of the reserves in the absence of claims, then the total reserve at epoch t is $A + ct - M(t)$, and negative reserve implies 'ruin'.

Corollary: We have

$$E\,\{M(t)\} = \lambda t\ E(X_i)$$

and

$$\operatorname{var}\{M(t)\} = \lambda t\ E(X_i^2). \qquad (3.5)$$

It can be shown that

$$\operatorname{cov}\{M(t), M(s)\} = \{\min\,(s, t)\}\,\{\lambda E(X_i^2)\}.$$

Compound Poisson process and linear combination of independent Poisson processes:

Consider Example 1(f).

The process $X(t) = aN_1(t) + bN_2(t)$ is a linear combination of two independent Poisson processes. The process can also be expressed as the compound Poisson process

$$X(t) = \sum_{i=1}^{N(t)} X_i,$$

where X_i, the amount of commission received from a subscription, is a random variable such that $\Pr\{X_i = a\} = \frac{2}{3}$ and $\Pr\{X_i = b\} = \frac{1}{3}$.

If follows that

$$E\{X(t)\} = E\{N(t)\}\, E\{X_i\} = (6t)\cdot\tfrac{1}{3}\cdot(2a + b) = 4at + 2bt$$

$$\operatorname{var}\{X(t)\} = 6t\, E\{X_i^2\} = 6t\,.\,\tfrac{1}{3}\,.\,(2a^2 + b^2) = 4a^2t + 2b^2t.$$

The two approaches are equivalent. The result may be stated in a more general form as follows:

Let $a_k > 0$, $k = 1, 2, \ldots, r\ (\geqslant 2)$ and $\Pr\{X_i = a_k\} = p_k$ for each i, $\Sigma\, p_k = 1$.

Then $X(t) = \sum_{i=1}^{N(t)} X_i$, where $\{N(t)\}$ is a Poisson process with parameter λ, is a compound Poisson process. For $t > 0$, let $\{N_i(t)\}$ be the number of jumps of value a_i for the process $\{X(t)\}$ which occur prior to t. Then we have

$$X(t) = a_1 N_1(t) + \ldots + a_r N_r(t),$$

where $\{N_k(t), t \geqslant 0\}$ is a Poisson process with parameter λp_k and $N_1(t), \ldots, N_r(t)$ are mutually independent. $N(t)$ is decomposed into r independent Poisson processes $N_k(t)$, $k = 1, 2, \ldots, r$.

Example 3(a). Customers arrive at a store in groups consisting of 1 or 2 individuals with equal probability and the arrival of groups is in accordance with a Poisson process with mean rate λ.

Here
$$p_k = \Pr\{X_i = k\} = \tfrac{1}{2} \text{ (for } k = 1, 2)$$
$$= 0 \text{ (otherwise)};$$

hence
$$P(s) = \sum_k p_k s^k = \tfrac{1}{2}s + \tfrac{1}{2}s^2$$

and so the generating function of $M(t)$, the total number of customers arriving in time t (by 3.4) is

$$G(s) = \exp\{\lambda t\,[\tfrac{1}{2}(s + s^2) - 1]\}.$$

The mean number of customers arriving in time t is $\frac{3}{2}\lambda t$. For $\lambda = \frac{1}{2}$ per minute and $t = 4$ minutes, the generating function will be

$$\exp[2\{\tfrac{1}{2}(s + s^2) - 1\}] = \{\exp(-2)\}\{\exp(s + s^2)\},$$

and the probability that the total number of arrivals is exactly 4 is

$$e^{-2}\left(\frac{1}{4!} + \frac{1}{2!} + \frac{1}{2!}\right) = 0.141.$$

Example 3(b). Suppose that the number of arrival epochs in an interval of length t is given by a Poisson process $\{N(t)\}$ with mean a and that the number of units arriving at an arrival epoch is given by a zero-truncated Poisson variate X_i, $i = 1, 2, \ldots$ with parameter λ. Then the total number $M(t)$ of units which arrive in an interval of length t is a Poisson cluster process with p.g.f.

$$G(P(s)) = \exp[at\{P(s) - 1\}],$$

where $P(s)$ is the p.g.f. of zero-truncated Poisson process (Exercise **1.5**), namely,

$$P(s) = (\exp\lambda - 1)^{-1}(\exp(\lambda s) - 1).$$

Hence

$$G(P(s)) = \exp\left[at\left\{\frac{(\exp(\lambda s) - 1)}{(\exp\lambda - 1)} - 1\right\}\right].$$

C. Pure Birth Process : Yule-Fury Process

In the classical Poisson process we assume that the conditional probabilities are constant. Here, the probability that k events occur between t and $t+h$ given that n events occurred by epoch t is given by

$$\begin{aligned} p_k(h) = \Pr\{N(h) = k \mid N(t) = n\} &= \lambda h + o(h), \quad k = 1 \\ &= o(h), \quad k \geqslant 2 \\ &= 1 - \lambda h + o(h), \; k = 0; \end{aligned}$$

$p_k(h)$ is independent of n as well as t. We can generalise the process by considering that λ is not a constant but is a function of n or t or both; the resulting processes will still be Markovian in character.

Here we consider that λ is a function of n, the population size at the instant. We assume that

$$\begin{aligned} p_k(h) = \Pr\{N(h) = k \mid N(t) = n\} &= \lambda_n h + o(h), \quad k = 1 \\ &= o(h), \quad k \geqslant 2 \qquad (3.6) \\ &= 1 - \lambda_n h + o(h), \; k = 0. \end{aligned}$$

Then proceeding as in Sec. 4.1, we shall have the following equation corresponding to (1.10)

$$p_n(t+h) = p_n(t)(1 - \lambda_n h) + p_{n-1}(t)\lambda_{n-1}h + o(h), \; n \geqslant 1.$$

Proceeding as before, we get

$$p'_n(t) = -\lambda_n p_n(t) + \lambda_{n-1}p_{n-1}(t), \quad n \geqslant 1 \qquad (3.7)$$

$$p'_0(t) = -\lambda_0 p_0(t). \qquad (3.8)$$

For given initial conditions, explicit expressions for $p_n(t)$ can be obtained from the above equations.

We shall consider here only a particular case of interest, the case $\lambda_n = n\lambda$ and describe a situation where this can happen.

While the above process is called *pure birth process*, the process corresponding to $\lambda_n = n\lambda$ is called *Yule-Furry process*.

Yule-Furry Process:

Consider a population whose members are either physical or biological entities. Suppose that members can give birth (by splitting or otherwise) to new members (who are exact replicas of themselves) but cannot die. We assume that in an interval of length h each member has a probability $\lambda h + o(h)$ of giving birth to a new member. Then, if n individuals are present at time t, the probability that there will be one birth between t and $t+h$ is $n\lambda h + o(h)$. If $N(t)$ denotes the total number of members by epoch t and $p_n(t) = \Pr\{N(t) = n\}$, then by putting $\lambda_n = n\lambda$ in (3.7) and (3.8), we obtain the following equations for $p_n(t)$:

$$p'_n(t) = -n\lambda p_n(t) + (n-1)\lambda p_{n-1}(t), \quad n \geqslant 1 \tag{3.9}$$

$$p_0'(t) = 0 \tag{3.10}$$

If the initial conditions are given, explicit expressions for $p_n(t)$ can be obtained.

Suppose that the initial condition is $p_1(0) = 1$, $p_i(0) = 0$ for $i \neq 1$, i.e. the process started with only one member at time $t = 0$. The solution can be obtained by the method of induction as follows :

For $n = 1$, we have

$$p_1'(t) = -\lambda p_1(t)$$

whose solution is

$$p_1(t) = c_1 e^{-\lambda t};$$

and putting $p_1(0) = 1$, we have $c_1 = 1$, i.e. $p_1(t) = e^{-\lambda t}$.

For $n=2$, we have

$$p_2'(t) = -2\lambda p_2(t) + \lambda p_1(t)$$

or

$$p_2'(t) + 2\lambda p_2(t) = \lambda p_1(t) = \lambda e^{-\lambda t}.$$

This linear equation has the integrating factor $e^{2\lambda t}$ and therefore

$$e^{2\lambda t} p_2(t) = \int \lambda e^{2\lambda t - \lambda t}\, dt = e^{\lambda t} + c_2\,;$$

since

$$p_2(0) = 0, \text{ we have } c_2 = -1$$

and

$$p_2(t) = e^{-2\lambda t}(e^{\lambda t} - 1) = e^{-\lambda t}(1 - e^{-\lambda t}).$$

Proceeding in this way, it can be shown that

$$p_n(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}, \; n \geqslant 1. \tag{3.11}$$

Solving (3.10) and noting that $p_0(0) = 0$, we get

$$p_0(t) = 0.$$

The distribution is of the geometric form (see Exercise 1.2). Its p.g.f. is given by

$$P(s,t) = \sum_{n=1}^{\infty} \{e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}\} s^n$$

$$= \frac{se^{-\lambda t}}{1 - s(1 - e^{-\lambda t})}. \tag{3.12}$$

The mean of the process is given by

$$E\{N(t)\} = e^{\lambda t};$$

Further, $\operatorname{var}\{N(t)\} = e^{\lambda t}(e^{\lambda t} - 1)$.

D. Time-dependent Poisson Processes

Here we assume that λ is a non-random function of time t, and that

$$\begin{aligned} \Pr\{N(h) = k \mid N(t) = n\} &= \lambda(t)h + o(h) \quad, k = 1 \\ &= o(h) \quad, k \geqslant 2 \qquad (3.13) \\ &= 1 - \lambda(t)h + o(h)\ , k = 0; \end{aligned}$$

then we get a time-dependent Poisson process. Here the postulate II of the classical Poisson process regarding homogeneity does not hold good and this process is called a non-homogeneous Poisson process. The equations of the process which can be obtained from those of Poisson process by replacing λ by $\lambda(t)$ are

$$p_n'(t) = -\lambda(t)p_n(t) + \lambda(t)p_{n-1}(t), \quad n \geqslant 1 \tag{3.14}$$

$$p_0'(t) = \lambda(t)p_0(t). \tag{3.15}$$

Theorem 4.7. The p.g.f. of a non-homogeneous process $\{N(t)\}$ is given by

$$Q(s, t) = \exp\{m(t)(s-1)\}, \tag{3.16}$$

where $\quad m(t) = \int_0^t \lambda(x)dx$ is the expectation of $N(t)$. (3.17)

Proof: Now $Q(s, t) = E\{s^{N(t)}\}$ with $Q(s, 0) = E\{s^{N(0)}\} = 1$.

Hence
$$\begin{aligned} Q(s, t+h) &= E\{s^{N(t+h)}\} \\ &= E\{s^{N(t+h)-N(t)}\, s^{N(t)}\} \\ &= E\{s^{N(t+h)-N(t)}\}\, E\{s^{N(t)}\} \\ &= E\{s^{N(t+h)-N(t)}\}\, Q(s, t), \end{aligned}$$

since $N(t)$ and $N(t+h) - N(t)$ are independent (by postulate I of Sec. 4.1 which still holds good; only postulate II does not).

Now
$$\begin{aligned} E\{s^{N(t+h)-N(t)}\} &= E\{s^{N(h)}\} \\ &= \sum_{k=0}^{\infty} \Pr\{N(h) = k\}s^k \\ &= \{1 - \lambda(t)h + o(h)\} + \{\lambda(t)h + o(h)\}\, s + o(h) \\ &= 1 + (s-1)\lambda(t)h + o(h). \end{aligned}$$

Hence we have

$$Q(s, t+h) = \{1 + (s-1)\lambda(t)h + o(h)\}\, Q(s, t)$$

or
$$\frac{Q(s, t+h) - Q(s, t)}{h} = (s-1)\lambda(t)\, Q(s, t) + \frac{o(h)}{h}.$$

We have, taking limits of both sides, as $h \to 0$,

$$\frac{\partial}{\partial t} Q(s,t) = (s-1)\lambda(t) Q(s,t).$$

Integrating we get

$$Q(s,t) = C \exp\{(s-1)\int_0^t \lambda(x)dx\}.$$

As $Q(s,0) = 0$, we have $C = 1$, so that

$$Q(s,t) = \exp\{m(t)(s-1)\},$$

where

$$m(t) = \int_0^t \lambda(x)dx.$$

Now

$$E\{N(t)\} = \left.\frac{\partial Q(s.t)}{\partial s}\right|_{s=1} = m(t),$$

so that $m(t)$ is the expectation of $N(t)$. ▲

Corollary: The probability of no occurrence in an interval of length t is given by

$$p_0(t) = Q(s,t)|_{s=0} = \exp\{-m(t)\} = \exp\{-\int_0^t \lambda(x)\,dx\}. \tag{3.18}$$

The probability of exactly k occurrences in an interval $(0,t)$ is given by

$$p_k(t) = \frac{1}{k!}\frac{\partial^k}{\partial s^k} Q(s,t)|_{s=0} = \frac{1}{k!}\{m(t)\}^k e^{-m(t)}$$

$$= \frac{1}{k!}\{\int_0^t \lambda(x)dx\}^k \exp\{-\int_0^t \lambda(x)dx\}, \quad k \geqslant 1. \tag{3.19}$$

That is, if we introduce $m(t)$ as *operational time* instead of absolute time t, the probability p_k of exactly k occurrences in the *operational time interval* $[0, m(t)]$ is given by the Poisson expression $\frac{[m(t)]^k \exp\{-m(t)\}}{k!}$. It can easily be verified that when $\lambda(t) = \lambda$, these lead to the corresponding probabilities of homogeneous Poisson process.

Further generalisation can be obtained by considering λ as a function of both n and t, i.e. $\lambda = \lambda_n(t)$.

The Polya process, which is such a generalisation, corresponds to

$$\lambda_n(t) = \frac{1+an}{1+at}.$$

E. Random Variation of the Parameter λ

In order to allow for lack of homogeneity and the influence of random factors on the development of the process, some random variation is introduced in the parameter λ of a Poisson process $N(t)$. These are considered below (see Cramér, 1969).

Case I : Suppose that λ is a random variable with probability density function $f(\lambda)$, $0 \leqslant \lambda \leqslant \infty$.

The probability of n occurrences in an interval of length t is then given by

$$p_n(t) = \Pr\{N(t) = n\} = \int_0^\infty \frac{(\lambda t)^n}{n!} e^{-\lambda t} f(\lambda) d\lambda.$$

The particular case when λ follows a gamma distribution is rather well known. Suppose that

$$f(\lambda) = \frac{h^k \lambda^{k-1} e^{-h\lambda}}{\Gamma(k)}$$

where k is a constant parameter. Then

$$p_n(t) = \binom{k+n-1}{n} p^k q^n, \; p = h/(t+h) = 1-q, \; n = 0, 1, \ldots$$

and $N(t)$ follows a negative binomial distribution with index k.

Case II Suppose that λ varies in a random way as time proceeds. Here, we consider λ to be a *random* function of time t, i.e. $\lambda = \lambda(t)$ is a stochastic process. A special case of Case II arises where the process $\lambda(t)$ is stationary. This has been studied as conditional Poisson process or 'doubly stochastic Poisson process'.

A conditional Poisson process is essentially a non-homogeneous Poisson process whose intensity function $\lambda(t)$ (or mean value function) is a stochastic process. These processes with strictly stationary intensity processes, have been studied by several authors (Bartlett, Kingman, McFadden, Serfozo etc., see Grandell (1976) for a detailed account).

F. Renewal Process

In the classical Poisson process, the intervals between successive occurrences are independently and identically distributed with a negative exponential distribution. Suppose that there is a sequence of events E such that the intervals between successive occurrences of E are distributed independently and identically but have a distribution not necessarily negative exponential; we have then a certain generalisation of the classical Poisson process: the corresponding process is called a *renewa process* (for details, see Ch. 6).

4.4 Birth and Death Process

In Sec. 4.3 C, we considered a pure birth process where

Pr {Number of births between t and $t+h$ is k,

given that the number of individuals at epoch t is n}

is given by

$$p(k, h \mid n, t) = \begin{cases} \lambda_n h + o(h), & k = 1 \\ o(h), & k \geqslant 2 \\ 1 - \lambda_n h + o(h), & k = 0. \end{cases} \qquad (4.1)$$

The above holds for all $n \geqslant 0$; λ_0 may or may not be equal to zero. Here k is a non-negative integer which implies that there can only be an increase by k, i.e. only births are considered possible. Now we suppose that there could also be a decrease by k, i.e. death (s) is also considered possible. In this case we shall further assume that

Pr {Number of deaths between t and $t + h$ is k,
given that the number of individuals at epoch t is n}

is given by

$$q(k, h \mid n, t) = \begin{cases} \mu_n h + o(h), & k = 1 \\ o(h), & k \geqslant 2 \\ 1 - \mu_n h + o(h), & k = 0. \end{cases} \tag{4.2}$$

The above holds for $n \geqslant k$; further $\mu_0 = 0$. With (4.1) and (4.2) we have, what is known as a *birth and death process*. Through a birth we have an increase by one and through a death we have a decrease by one in the number of "individuals". The probability of more than one birth or more than one death in an interval of length h is $o(h)$. Let $N(t)$ denote the total number of individuals at epoch t starting from $t = 0$ and let $p_n(t) = \Pr\{N(t) = n\}$. Consider the interval between 0 and $t + h$; suppose that it is split into two periods $(0, t)$ and $(t, t + h)$. The event $\{N(t + h) = n, n \geqslant 1\}$, (having probability $p_n(t + h)$) can occur in a number of mutually exclusive ways.

These would include events involving more than one birth and/or more than one death between t and $t + h$. By our assumption, the probability of such an event is $o(h)$. There will remain four other events to be considered:

$A_{ij}: (n - i + j)$ individuals by epoch t, i birth and j death between t and $t + h$, $i, j = 0, 1$.

We have

$$\begin{aligned} \Pr(A_{00}) &= p_n(t)\{1 - \lambda_n h + o(h)\}\{1 - \mu_n h + o(h)\} \\ &= p_n(t)\{1 - (\lambda_n + \mu_n)h + o(h)\}; \\ \Pr(A_{10}) &= p_{n-1}(t)\{\lambda_{n-1} h + o(h)\}\{1 - \mu_{n-1} h + o(h)\} \\ &= p_{n-1}(t)\,\lambda_{n-1} h + o(h); \\ \Pr(A_{01}) &= p_{n+1}(t)\{1 - \lambda_{n+1} h + o(h)\}\{\mu_{n+1} h + o(h)\} \\ &= p_{n+1}(t)\,\mu_{n+1} h + o(h); \\ \Pr(A_{11}) &= p_n(t)\{\lambda_n h + o(h)\}\{\mu_n h + o(h)\} \\ &= o(h). \end{aligned}$$

Hence we have, for $n \geqslant 1$

$$\begin{aligned} p_n(t + h) = p_n(t)\{1 - (\lambda_n + \mu_n)h\} &+ p_{n-1}(t)\,\lambda_{n-1} h \\ &+ p_{n+1}(t)\,\mu_{n+1} h + o(h) \end{aligned} \tag{4.3}$$

or
$$\frac{p_n(t+h)-p_n(t)}{h} = -(\lambda_n+\mu_n)\,p_n(t)+\lambda_{n-1}\,p_{n-1}(t) + \mu_{n+1}\,p_{n+1}(t)+\frac{o(h)}{h},$$

and taking limits, as $h \to 0$, we have

$$p'_n(t) = -(\lambda_n+\mu_n)\,p_n(t)+\lambda_{n-1}\,p_{n-1}(t)+\mu_{n+1}\,p_{n+1}(t), \quad n \geqslant 1 \qquad (4.4)$$

For $n = 0$, we have

$$p_0(t+h) = p_0(t)\,\{1-\lambda_0 h+o(h)\}+p_1(t)\,\{1-\lambda_0 h+o(h)\}\,\{\mu_1 h+o(h)\} \qquad (4.5)$$

$$= p_0(t)-\lambda_0 h\,p_0(t)+\mu_1 h\,p_1(t)$$

or
$$\frac{p_0(t+h)-p_0(t)}{h} = -\lambda_0\,p_0(t)+\mu_1\,p_1(t)+\frac{o(h)}{h}$$

whence we get

$$p'_0(t) = -\lambda_0\,p_0(t)+\mu_1\,p_1(t). \qquad (4.6)$$

If at time $t = 0$, there were i ($\geqslant 0$) individuals, then the initial condition is

$$p_n(0) = 0,\ n \neq i\ ;\ p_i(0) = 1. \qquad (4.7)$$

The equations (4.4) and (4.6) are the equations of the birth and death process. The birth and death processes play an important role in queueing theory. They also have interesting applications in diverse other fields such as economics, biology, ecology, reliability theory etc.

Note: The result about existence of solutions of (4.4) and (4.6) is stated below without proof.

For arbitrary $\lambda_n \geqslant 0$, $\mu_n \geqslant 0$, there always exists a solution $p_n(t)$ ($\geqslant 0$) such that $\Sigma\,p_n(t) \leqslant 1$. If λ_n, μ_n are bounded, the solution is unique and satisfies $\Sigma\,p_n(t) = 1$.

Birth and Death Rates

Some particular values of λ_n and μ_n are of special interest.

When $\lambda_n = \lambda$, i.e. λ_n is independent of the population size n, then the increase may be thought of as due to an external source such as *immigration*.

When $\lambda_n = n\lambda$, we have the case of (linear) birth; $\lambda_n h = n\lambda h$ may be considered as the probability of one birth in an interval of length h given that n individuals are present (at the instant from which the interval commences), the probability of one individual giving a birth being λh. (i.e. rate of birth in unit internal is λ per individual). Here $\lambda_0 = 0$.

When $\mu_n = \mu$, the decrease may be thought of as due to a factor such as *emigration*.

When $\mu_n = n\mu$, we have the case of (linear) death, the rate of death in unit interval being μ per individual.

Particular Cases

I. Immigration-Emigration Process

For $\lambda_n = \lambda$ and $\mu_n = \mu$ we have what is known as *immigration-emigration* process. The process associated with the simple queueing model $M/M/1$ (which we discuss in Sec. 10.2) is such a process.

II. Linear Growth Process

(a) *Generating function* : In the Yule-Furry process (Sec. 4.3 B) one is concerned with a population whose members can give birth only (to new members) but cannot die. Let us consider the case where both births and deaths can occur. Suppose that the probability that a member gives birth to a new member in a small interval of length h is $\lambda h + o(h)$ and the probability that a member dies is $\mu h + o(h)$. Then, if n members are present at the instant t, the probability of one birth between t and $t + h$ is $n\lambda h + o(h)$ and that of one death is $n\mu h + o(h)$, $n \geqslant 1$.

We have thus a birth and death process with

$$\lambda_n = n\lambda,\ \mu_n = n\mu\ (n \geqslant 1),\ \lambda_0 = \mu_0 = 0$$

If $X(t)$ denotes the total number of members at time t, then from (4.4) and (4.6) we have the following differential-difference equations for $p_n(t) = \Pr\{X(t) = n\}$:

$$p_n'(t) = -n(\lambda + \mu)p_n(t) + \lambda(n-1)p_{n-1}(t) + \mu(n+1)p_{n+1}(t), n \geqslant 1 \quad (4.8)$$

$$p_0'(t) = \mu p_1(t). \quad (4.9)$$

If the initial population size is i, i.e. $X(0) = i$, then we have the initial condition $p_i(0) = 1$ and $p_n(0) = 0$, $n \neq i$.

Let

$$P(s, t) = \sum_{n=0}^{\infty} p_n(t)s^n \qquad \text{be the p.g.f. of } \{p_n(t)\}.$$

Then

$$\frac{\partial P}{\partial s} = \sum_{n=1}^{\infty} np_n(t)s^{n-1} \text{ and } \frac{\partial P}{\partial t} = \sum_{n=0}^{\infty} p_n'(t)s^n.$$

Multiplying (4.8) by s^n and adding over $n = 1, 2, 3 \ldots$ and adding (4.9) thereto, we get

$$\frac{\partial P}{\partial t} = -(\lambda + \mu)\sum_{n=1}^{\infty} np_n(t)s^n + \lambda \sum_{n=1}^{\infty}(n-1)p_{n-1}(t)s^n$$

$$+ \mu\left\{\sum_{n=1}^{\infty}(n+1)p_{n+1}(t)s^n + p_1(t)\right\}$$

$$= -(\lambda + \mu)s\frac{\partial P}{\partial s} + \lambda s^2 \frac{\partial P}{\partial s} + \mu\frac{\partial P}{\partial s}$$

$$= \{\mu - (\lambda + \mu)s + \lambda s^2\}\frac{\partial P}{\partial s}. \quad (4.10)$$

$P(s, t)$ thus satisfies a partial differential equation of Lagrangian type. We shall not discuss here the method of solution; the solution with the initial condition $X(0) = i$, is given by

$$
\begin{aligned}
P(s,t) &= \left[\frac{\mu(1-s)-(\mu-\lambda s)e^{-(\lambda-\mu)t}}{\lambda(1-s)-(\mu-\lambda s)e^{-(\lambda-\mu)t}}\right]^i \\
&= \left[\frac{\mu\{1-e^{-(\lambda-\mu)t}\}-\{\mu-\lambda e^{-(\lambda-\mu)t}\}s}{\{\lambda-\mu e^{-(\lambda-\mu)t}-\lambda\{1-e^{-(\lambda-\mu)t}\}s}\right]^i. \qquad (4.11)
\end{aligned}
$$

Explicit expression for $p_n(t)$ can be obtained from the above by expanding $P(s, t)$ as a power series in s.

(b) *Mean population size*: We can obtain the mean population size by differentiating $P(s, t)$ partially with respect to s and putting $s = 1$. It can however be obtained directly from (4.8) and (4.9) without obtaining $P(s, t)$ as follows :

Let
$$E\{X(t)\} = M(t) = \sum_{n=1}^{\infty} np_n(t)$$

and
$$E\{X^2(t)\} = M_2(t) = \sum_{n=1}^{\infty} n^2 p_n(t).$$

Multiplying both sides of (4.8) by n and adding over for $n = 1, 2, 3, \ldots$, we have

$$
\begin{aligned}
\sum_{n=1}^{\infty} n p_n'(t) &= -(\lambda+\mu)\sum_{n=1}^{\infty} n^2 p_n(t) + \lambda \sum_{n=1}^{\infty} n(n-1)p_{n-1}(t) \\
&\quad + \mu \sum_{n=1}^{\infty} n(n+1)p_{n+1}(t). \qquad (4.12)
\end{aligned}
$$

Now

$$
\begin{aligned}
\sum_{n=1}^{\infty} n(n-1)\,p_{n-1}(t) &= \sum_{n=1}^{\infty} (n-1)^2 p_{n-1}(t) + \sum_{n=1}^{\infty} (n-1)\,p_{n-1}(t) \\
&= M_2(t) + M(t); \\
\sum_{n=1}^{\infty} n(n+1)\,p_{n+1}(t) &= \sum_{n=1}^{\infty} (n+1)^2 p_{n+1}(t) - \sum_{n=1}^{\infty} (n+1)\,p_{n+1}(t) \\
&= \{M_2(t) - p_1(t)\} - \{M(t) - p_1(t)\} \\
&= M_2(t) - M(t);
\end{aligned}
$$

and
$$\sum_{n=1}^{\infty} np'(t) = M'(t).$$

Hence from (4.12) we get

$$
\begin{aligned}
M'(t) &= -(\lambda+\mu)\,M_2(t) + \lambda\,\{M_2(t) + M(t)\} + \mu\,\{M_2(t) - M(t)\} \\
&= (\lambda-\mu)\,M(t).
\end{aligned}
$$

The solution of the above differential equation (which $M(t)$ satisfies) is easily found to be

$$M(t) = Ce^{(\lambda-\mu)t}.$$

The initial condition gives $M(0) = \sum_{n=1}^{\infty} np_n(0) = i$, whence $C = M(0) = i$.

We have therefore,

$$M(t) = ie^{(\lambda-\mu)t}. \tag{4.13}$$

The second moment $M_2(t)$ of $X(t)$ can also be calculated in the same way.

Limiting case : As $t \to \infty$, the mean population size $M(t)$ tends to 0 for $\lambda < \mu$ (birth rate smaller than death rate) or to ∞ for $\lambda > \mu$ (birth rate greater than death rate) and to the constant value i when $\lambda = \mu$.

(c) *Extinction probability:* Since $\lambda_0 = 0$, 0 is an absorbing state, i.e. once the population size reaches 0, it remains at 0 thereafter This is the interesting case of extinction of the population. We can determine the probability of extinction as follows:

Suppose, for simplicity that $X(0) = 1$, i.e. the process starts with only one member at time 0. Then from (4.11) we can write $P(s, t)$ as

$$P(s, t) = \frac{a - bs}{c - ds} = \frac{a}{c} \cdot \frac{1 - bs/a}{1 - ds/c}$$

where $\qquad a = \mu\{1 - e^{-(\lambda-\mu)t}\}$

and $\qquad c = \lambda - \mu e^{-(\lambda-\mu)t}$.

$\Pr\{X(t) = 0\} = p_0(t)$, the constant term in the expansion of $P(s, t)$ as a power series in s, is given by

$$\frac{a}{c} = \frac{\mu\{1 - e^{-(\lambda-\mu t)}\}}{\lambda - \mu e^{-(\lambda-\mu)t}}. \tag{4.14}$$

The probability that the population will eventually die out is given by $\lim p_0(t)$ as $t \to \infty$ and can be obtained from the above by letting $t \to \infty$.

If $\lambda > \mu$, then

$$\begin{aligned}\lim_{t\to\infty} p_0(t) &= \lim_{t\to\infty} \frac{\mu\{1 - e^{-(\lambda-\mu)t}\}}{\lambda - \mu e^{-(\lambda-\mu)t}} \\ &= \frac{\mu}{\lambda} < 1.\end{aligned}$$

If $\lambda < \mu$, then

$$\lim_{t\to\infty} p_0(t) = \lim_{t\to\infty} \frac{\mu\{1 - e^{-(\mu-\lambda)t}\}}{\lambda - \mu e^{-(\mu-\lambda)t}} = 1. \tag{4.15}$$

and $\qquad \lim_{t\to\infty} p_n(t) = 0$ for $n \neq 0$

In other words, the probability of ultimate extinction is 1 when $\mu > \lambda$ (i.e. when the death rate is greater than the birth rate) and is μ/λ when $\mu < \lambda$.

III. Linear Growth with Immigration

In II, we have $\lambda_0 = 0$ and, as a result, if the population size reaches zero at any time, it remains at zero thereafter. Here 0 is an absorbing state. If we consider $\lambda_n = n\lambda + a \ (a > 0)$, $\mu_n = n\mu \ (n \geqslant 0)$ we get what is known as a linear growth process with immigration, where 0 is not an absorbing state.

IV. Immigration-Death Process

If $\lambda_n = \lambda$ and $\mu_n = n\mu$, we get what is known as an immigration-death process. This corresponds to the Markovian queue with infinite number of channels, i.e. the queue $M/M/\infty$.

V. Pure Death Process

Here $\lambda_n = 0$ for all n, i.e. an individual cannot give birth to a new individual and the probability of a death of an individual in $(t, t+h)$ is $\mu h + o(h)$. Then, if n individuals are present at time t, the probability of one death in $(t, t+h)$ is $n\mu h + o(h)$.

The birth and death process is a special type of continuous time Markov process with discrete state space 0, 1, 2, ... such that the probability of transition from state i to state j in Δt is $o(\Delta t)$ whenever $|i-j| \geqslant 2$. In other words, changes take place through transitions only from a state to its immediate neighbouring state.

In the next section we consider certain aspects of more general continuous time Markov processes with discrete space. Such processes are also known as *continuous time Markov chains.*

4.5 Markov Processes with Discrete State Space (Continuous Time Markov Chains)

A. Introduction

Let $X(t)$ be a continuous parameter Markov Process with state space $N = \{0, 1, 2, \ldots, i, \ldots, j, \ldots\}$.

Further let $X(t)$ be time-homogeneous or stationary; then the (transition) probability of a transition from state i to state j during the time interval from epoch T to epoch $T+t$ does not depend on the initial time T but depends only on the elapsed time t and on the initial and terminal states i and j. We can thus write

$$\Pr\{X(T+t) = j \mid X(T) = i\} = p_{ij}(t), \qquad i, j = 0, 1, 2, \ldots, t \geqslant 0. \tag{5.1}$$

In particular, we have

$$\Pr\{X(t) = j \mid X(0) = i\} = p_{ij}(t).$$

We have

$$0 \leqslant p_{ij}(t) \leqslant 1 \text{ for each } i, j, t$$

and

$$\sum_j p_{ij}(t) = 1.$$

Let $p_j(t) = \Pr\{X(t) = j\}$ be the state probability at epoch t; then

$$\begin{aligned} p_j(t) &= \Pr\{X(t) = j\} \\ &= \sum_i \Pr\{X(t) = j \text{ and } X(0) = i\} \\ &= \sum_i \Pr\{X(0) = i\} \Pr\{X(t) = j \mid X(0) = i\} \\ &= \sum_i \Pr\{X(0) = i\}\, p_{ij}(t). \end{aligned} \tag{5.2}$$

We have $\sum_j p_j(t) = 1$ for each $t \geqslant 0$.

Let us denote the transition probability matrix by

$$P(t) = (p_{ij}(t)). \tag{5.3}$$

Setting $p_{ij}(0) = \delta_{ij}$, we get

$$P(0) = I. \tag{5.4}$$

We shall assume here that the functions $p_{ij}(t)$ are continuous and differentiable for $t \geqslant 0$. (For details, see Chung).

The waiting time for a change of state: Suppose that $X(t)$ is a homogeneous Markov process and that at time $t_0 = 0$ the state of the process $X(t_0) = X(0) = i$ is known. The time taken for a change of state from state i is a random variable, say τ, which is called the waiting time for a change of state from state i.

We have

$$\begin{aligned} &\Pr\{\tau > s + t \mid X(0) = i\} \\ &\qquad = \Pr\{\tau > s + t \mid X(0) = i, \tau > s\} \Pr\{\tau > s \mid X(0) = i\} \\ &\qquad = \Pr\{\tau > s + t \mid X(s) = i\} \Pr\{\tau > s \mid X(0) = i\}. \end{aligned}$$

If we denote $F(t) = \Pr\{\tau > t \mid X(0) = i\}$, $t > 0$ then the above can be written as

$$F(s + t) = F(s)\, F(t), \text{ for } s, t > 0,$$

which is satisfied iff $F(t)$ is of the form

$$F(t) = e^{-\lambda t}, \ t > 0, \lambda > 0.$$

Thus the waiting time τ has exponential distribution with parameter λ, which is called the transition density from state i.

B. Chapman-Kolmogorov Equations

The transition probability $p_{ij}(t + T)$ is the probability that given that the state was i at epoch 0, it is state j at epoch $t + T$; but in passing from state i to state j in time $(t + T)$ the process passes through some state k in time t. Thus

$$p_{ij}(t + T) = \sum_k \Pr\{X(t + T) = j,\ X(t) = k \mid X(0) = i\}$$

$$= \sum_k \Pr\{X(0) = i,\ X(t) = k,\ X(t+T) = j\}/\Pr\{X(0) = i\}$$

$$= \sum_k \frac{\Pr\{X(0)=i, X(t)=k\}}{\Pr\{X(0)=i\}} \times \frac{\Pr\{X(0)=i, X(t)=k, X(t+T)=j\}}{\Pr\{X(0)=i, X(t)=k\}}$$

$$= \sum_k \Pr\{X(t) = k \mid X(0) = i\} \Pr\{X(t+T) = j \mid X(0) = i, X(t) = k\}.$$

Since $X(t)$ is a Markov process,

$$\begin{aligned}\Pr\{X(t+T) = j \mid X(0) = i,\ X(t) = k\} &= \Pr\{X(t+T) = j \mid X(t) = k\}\\ &= p_{kj}(T).\end{aligned}$$

Thus we get the relation

$$p_{ij}(t+T) = \sum_k p_{ik}(t) p_{kj}(T), \tag{5.5}$$

which holds for all states i, j and $t \geqslant 0$, $T \geqslant 0$; it is called the Chapman-Kolmogorov equation. In matrix notation it can be written as

$$P(t+T) = P(t)P(T). \tag{5.6}$$

The equations (5.5) and (5.6) correspond to (2.4) and (2.5) (of Ch. 3) respectively which hold for discrete space Markov chain.

Denote the right-hand derivative at zero by

$$a_{ij} = \frac{d}{dt} p_{ij}(t)|_{t=0};\ i \neq j \tag{5.7}$$

$$a_{ii} = \frac{d}{dt} p_{ii}(t)|_{t=0}.$$

Then

$$a_{ij} = \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t) - p_{ij}(0)}{\Delta t} = \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t)}{\Delta t}$$

or

$$p_{ij}(\Delta t) = a_{ij}\Delta t + o(\Delta t), \quad i \neq j \tag{5.8}$$

and

$$a_{ii} = \lim_{\Delta t \to 0} \frac{p_{ii}(\Delta t) - p_{ii}(0)}{\Delta t} = \lim_{\Delta t \to 0} \frac{p_{ii}(\Delta t) - 1}{\Delta t}$$

or

$$p_{ii}(\Delta t) = 1 + a_{ii}\Delta t + o(\Delta t). \tag{5.9}$$

It can be seen from the above relations that $a_{ij} \geqslant 0$, $i \neq j$ and $a_{ii} < 0$.

From $\sum_j p_{ij}(t) = 1$, using (5.7) we get

$$\sum_j a_{ij} = 0$$

or

$$\sum_{i \neq i} a_{ij} = -a_{ii}. \tag{5.10}$$

The quantities a_{ij} are called *transition densities* and the matrix

$$A = (a_{ij})$$

is called the *transition density matrix* or *rate matrix* of the process. The matrix is such that

(1) its off-diagonal elements are non-negative and the diagonal elements are negative;

(2) the sum of elements of each row is zero, the sum of the off-diagonal elements being equal in magnitude but opposite in sign to the diagonal elements.

Differentiating (5.5) with respects to T, we get

$$p'_{ij}(t+T) = \frac{\partial}{\partial T} p_{ij}(t+T) = \sum_k p_{ik}(t) \frac{d}{dT} p_{kj}(T).$$

Putting $T = 0$, we get

$$p'_{ij}(t) = \sum_k p_{ik}(t) a_{kj}. \tag{5.11}$$

Or, in matrix notation

$$P'(t) = P(t)A. \tag{5.11a}$$

Similarly we can get

$$\frac{d}{dT} p_{ij}(T) = \sum_k a_{ik} p_{kj}(T).$$

Replacing T by t, we can write this as

$$p'_{ij}(t) = \sum_k a_{ik} p_{kj}\, t) \tag{5.12}$$

or

$$P'(t) = AP(t). \tag{5.12a}$$

Equations (5.11) and (5.12) which give Chapman-Kolmogorov equations as differential equations are called respectively Forward and Backward Kolmogorov equations.

Solution of the Equations for a Finite State Process

When the rate matrix is given, the equations (5.11) or (5.11a) together with the initial conditions $p_{ij} = \delta_{ij}$ (or $P(0) = I$) yield as solution the unknown probabilities $p_{ij}(t)$. We consider below a method of solution for a process with finite number of states. From (5.11a) we see at once that the solutions can be written in the form

$$P(t) = e^{At}, \tag{5.13}$$

where the matrix

$$e^{At} = I + \sum_{n=1}^{\infty} \frac{A^n t^n}{n!}. \tag{5.13a}$$

Assume that the eigenvalues of A are all distinct. Then from the spectral theorem of matrices (see Sec. 1.7 or 3.4), we have

$$A = HDH^{-1},$$

where H is a non-singular matrix (formed with the right eigenvectors of A) and D is the diagonal matrix having for its diagonal elements the eigenvalues of A. Now, 0 is an eigenvalue of A and if $d_i \neq 0$, $i = 1, \ldots, m$ are the other distinct eigenvalues, then

$$D = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & d_1 & \dots & 0 \\ \dots & & \dots & \dots \\ \dots & & \dots & \dots \\ 0 & 0 & \dots & d_m \end{pmatrix}.$$

We then have

$$D^n = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & d_1^n & \dots & 0 \\ \dots & & \dots & \dots \\ \dots & & \dots & \dots \\ 0 & 0 & \dots & d_m^n \end{pmatrix}$$

and $$A^n = HD^nH^{-1}.$$

Substituting in (5.13), we get

$$\begin{aligned} P(t) &= I + \sum_{n=1}^{\infty} \frac{(HD^nH^{-1})\, t^n}{n!} \\ &= H\left\{I + \sum_{n=1}^{\infty} \frac{D^n t^n}{n!}\right\} H^{-1} \\ &= H\, e^{Dt}\, H^{-1}, \qquad (5.14) \end{aligned}$$

where

$$e^{Dt} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & e^{d_1 t} & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & e^{d_m t} \end{pmatrix}.$$

The right-hand side of (5.14) gives explicit solution of the matrix P(t). Note that even in the general case when the eigenvalues of A are not necessarily distinct, a canonical representation of $A = SZS^{-1}$ exists (sec. 3.4). Using this, $P(t)$ can be obtained in a modified form.

Example 5(a). Poisson process: If events occur in accordance with a Poisson process $N(t)$ with mean λt, then

$$\begin{aligned} p_{i,i+1}(\Delta t) &= \Pr \{\text{the process goes to state } i+1 \text{ from state } i \text{ in time } \Delta t\} \\ &= \Pr \{\text{one event occurs in time } \Delta t\} \\ &= \Pr \{N(\Delta t) = 1\} \\ &= \lambda \Delta t + o(\Delta t), \\ p_{i,i}(\Delta t) &= 1 - \lambda \Delta t + o(\Delta t) \end{aligned}$$

and $$p_{i,j}(\Delta t) = o(\Delta t),\ j \neq i,\ i+1.$$

By comparing with (5.8) and (5.9), we have

$$a_{i,i+1} = \lambda,\ a_{i,i} = -\lambda,\ a_{i,j} = 0 \text{ for } j \neq i,\ i+1.$$

The Kolmogorov forward equations become

$$\begin{aligned} p'_{i,i}(t) &= -\lambda p_{i,i}(t) \\ p_{i,j}(t) &= -\lambda p_{i,j}(t) + \lambda p_{i,j-1}(t), \; j = i+1, i+2, \ldots \end{aligned} \tag{5.15}$$

Let $p_j(t) = \Pr\{N(t) = j\}$ and $p_0(0) = 0$, $p_n(0) = 0$, $n \neq 0$. Using (5.2) we get $p_j(t) \equiv p_{0j}(t)$, $j = 0, 1, 2, \ldots$ Thus (5.15) become identical with (1.11) and (1.12) obtained directly in terms of $p_j(t)$.

Example 5(b). Two-state process: Suppose that a certain system can be considered to be in two states: "Operating" and "Under repair" (denoted by 1 and 0 respectively). Suppose that the lengths of operating period and the period under repair are independent random variables having negative exponential distributions with means $1/b$ and $1/a$ respectively ($a, b > 0$). The evolution of the system can be described by a Markov process with two states 0 and 1.

Now,

$$\begin{aligned} p_{01}(\Delta t) &= \Pr\{\text{change of state from 0 to 1 in time } \Delta t\} \\ &= \Pr\{\text{repair being completed in time } \Delta t\} \\ &= a\Delta t + o(\Delta t) \end{aligned}$$

and

$$\begin{aligned} p_{10}(\Delta t) &= \Pr\{\text{change of state from 1 to 0 in } \Delta t\} \\ &= b\Delta t + o(\Delta t) \end{aligned}$$

Thus the transition densities are

$$a_{01} = a, \qquad a_{10} = b$$

and

$$a_{00} = -a, \quad a_{11} = -b$$

so that

$$A = \begin{pmatrix} -a & a \\ b & -b \end{pmatrix}.$$

The Kolmogorov forward equations, for $i = 0, 1$, are

$$\begin{aligned} p'_{i0}(t) &= -a\, p_{i0}(t) + b\, p_{i1}(t) \\ p'_{i1}(t) &= \quad a\, p_{i0}(t) - b\, p_{i1}(t). \end{aligned}$$

Now we proceed to find the transition probabilities $p_{ij}(t)$.

Using

$$p_{00}(t) + p_{01}(t) = 1, \; p_{10}(t) + p_{11}(t) = 1,$$

we get

$$p'_{00}(t) - (a+b)\, p_{00}(t) = b$$

and

$$p'_{11}(t) + (a+b)\, p_{11}(t) = a.$$

The solution of the first of these differential equations is

$$p_{00}(t) = \frac{b}{a+b} + Ce^{-(a+b)t}.$$

With $p_{00}(0) = 1$, we get $C = \dfrac{a}{a+b}$, so that

$$p_{00}(t) = \frac{b}{a+b} + \frac{a}{a+b} e^{-(a+b)t}.$$

Hence $$p_{01}(t) = 1 - p_{00}(t) = \frac{a}{a+b} - \frac{a}{a+b} e^{-(a+b)t}.$$

Similarly, the solution of the second differential equation with the initial condition $p_{11}(0) = 1$, gives

$$p_{11}(t) = \frac{a}{a+b} + \frac{b}{a+b} e^{-(a+b)t}$$

and hence $$\begin{aligned} p_{10}(t) &= 1 - p_{11}(t) \\ &= \frac{b}{a+b} - \frac{b}{a+b} e^{-(a+b)t}. \end{aligned}$$

Let $p_j(t)$ be the probability that the system is in state j at time t, $j = 0, 1$, and let $p_0(0) = 1$, $p_n(0) = 0$, $n \neq 0$. Then $p_j(t) \equiv p_{0j}(t)$, $j = 0, 1$.

Alternative method: This is considered to show how to proceed with the matrix or algebraic method of solution outlined above in case the Kolmogorov differential equations are not easily solvable. In the example under consideration, it is not the case. However, the solution given below is expected to serve as an illustration of the method.

In matrix notation, the Kolmogorov equations can be put as

$$P'(t) = (p'_{ij}(t)) = P(t)\, A,$$

whose solution (as given in 5.14) is

$$P(t) = He^{Dt}H^{-1}, \tag{5.16}$$

where $$A = HDH^{-1}.$$

The matrix A has two distinct eigenvalues 0 and $-(a+b)$.
The right eigenvectors corresponding to the eigenvalues 0 and $-(a+b)$, are the column vectors $(1, 1)'$ and $(a, -b)'$ respectively. The matrix H which is formed by these two eigenvectors is

$$H = \begin{pmatrix} 1 & a \\ 1 & -b \end{pmatrix}$$

so that $$H^{-1} = \begin{pmatrix} b/(a+b) & a/(a+b) \\ 1/(a+b) & -1/(a+b) \end{pmatrix}.$$

The diagonal matrix D formed with the eigenvalues as diagonal elements is

$$D = \begin{pmatrix} 0 & 0 \\ 0 & -(a+b) \end{pmatrix}$$

and $$e^{Dt} = \begin{pmatrix} 1 & 0 \\ 0 & e^{-(a+b)t} \end{pmatrix}.$$

Substituting these values in (5.16), we get $p_{ij}(t)$ as obtained above.

C. Limiting Distribution (Ergodicity of homogeneous Markov process)

We recall the result on limiting distribution of certain types of Markov chains. Theorem 3.6 states: For an irreducible ergodic Markov chain, the limits

$$\lim_{n\to\infty} p_{jk}^{(n)} = v_k$$

exist and form a probability distribution. Further, the limits v_k can be obtained as unique solutions of the equation

$$V(P - I) = 0. \tag{5.17}$$

where $$V = (v_0, \ldots, v_k, \ldots), \; P = (p_{ij}).$$

A similar elegant result holds for continuous parameter Markov processes as well. We shall state the result without proof. Here similar definitions for the classification of the states (which constitute a discrete state-space) will be used.

Theorem 4.8. Suppose that the Markov process $\{X(t)\}$ is time-homogeneous, irreducible and ergodic; also that its t.p.m. is $P(t) = (p_{ij}(t))$, $i, j = 0, 1, 2, \ldots$ and the matrix of transition densities (or rate matrix) is

$$A = (a_{ij}),$$

where $$a_{ij} = p'_{ij}(t)\,|_{t=0}.$$

Then given any state j,

$$\lim_{t\to\infty} p_{ij}(t) = v_j \tag{5.18}$$

exists and is the same for all initial states $i = 0, 1, 2, \ldots$. The asymptotic values v_j respresent a probability distribution, i.e.

$$0 \leqslant v_j \leqslant 1, \; \sum_j v_j = 1.$$

The values v_j can then be determined as solutions of the system of linear equations

$$\sum_i v_i a_{ij} = 0, \quad j = 0, 1, 2, \ldots \tag{5.19}$$

or in matrix notation,

$$VA = \mathbf{0}, \; V = (v_0, v_1, \ldots) \tag{5.19a}$$

by using the normalising condition

$$\sum_j v_j = 1.$$

Note 1: The eq. (5.19) can be obtained from the forward Kolmogorov eq. (5.11) by putting

$$\lim_{t\to\infty} p_{ij}(t) = v_j \quad \text{and} \quad \lim_{t\to\infty} p'_{ij}(t) = 0.$$

Note 2: The eqs. (5.17) and (5.19a) for discrete and continuous parameter processes respectively are similar in structure. The matrices $(P - I)$ and A both have non-negative off-diagonal elements, strictly negative diagonal elements and zero row sums. If the number of states are finite, say, m, then both $(P-I)$ and A are of rank $(m-1)$. Then V can be easily determined from any of the $(m-1)$ equations (out of m equations contained in the relations (5.17) or (5.19a)) and the normalising condition $\sum_j v_j = 1$.

Example 5 (c). Consider the two-state process given in Example 5(b). Here

$$v_0 = \lim_{t\to\infty} p_{00}(t) = \lim_{t\to\infty} p_{10}(t) = b/(a+b)$$

and

$$v_1 = \lim_{t\to\infty} p_{01}(t) = \lim_{t\to\infty} p_{11}(t) = a/(a+b).$$

These limiting probabilities can also be obtained from the equations (5.19) which become

$$-a\,v_0 + b\,v_1 = 0$$

$$a\,v_0 - b\,v_1 = 0.$$

Here there are two states and so the matrix A is of rank 1 and the two equations are equivalent. Each of them yields $v_1 = (a/b)\,v_0$. Using the normalising condition $v_0 + v_1 = 1$, we get

$$v_0 = b/(a+b),\ v_1 = a/(a+b).$$

When the number of states is finite and when only the limiting probabilities v_j are needed it is easier and more convenient to determine them from (5.19) or (5.19a).

Example 5(d). Suppose that in a queueing system which has m service channels, the demand for service arises in accordance with a Poisson process with parameter a. Suppose that the servicing time in each channel is exponential with parameter b. Further suppose that there is no storage facility, in the sense that a demand which arrives at the moment when any channel is free is received and is processed, whereas a demand which is received when all the m channels are busy is rejected and leaves the system. This is called a *loss system*.

Suppose that state k implies that k channels are busy. Then we get a Markov process with $(m+1)$ states $0, 1, \ldots, m$. The transition probabilities are given by

$$p_{i,i+1}(\Delta t) = \Pr\{\text{one demand is received for processing in time } \Delta t\}$$

$$= a\,\Delta t + o(\Delta t),\ 0 \leqslant i < m$$

$$p_{i,j}(\Delta t) = o(\Delta t),\ j > i+1.$$

Pr {one service demand is met in time Δt}

$$= b\,\Delta t + o(\Delta t),$$

and so if i channels are working

Pr {one service demand is met in time Δt}

$$= i\,b\,\Delta t + o(\Delta t).$$

i.e.
$$p_{i,i-1}(\Delta t) = i\,b\,\Delta t + o(\Delta t)$$
and
$$p_{i,j}(\Delta t) = o(\Delta t),\ j < i-1.$$
Thus
$$a_{01} = a,\ a_{00} = -a$$
$$\left.\begin{aligned} a_{ij} &= ib &&, j = i-1 \\ &= a &&, j = i+1 \\ &= -(a+ib) &&, j = i \end{aligned}\right\} 1 \leqslant i < m$$
$$a_{m,m-1} = mb,\ a_{m,m} = -mb$$
$$A = \begin{pmatrix} -a & a & 0 & 0 & \dots & 0 & 0 \\ b & -(a+b) & a & 0 & \dots & 0 & 0 \\ 0 & 2b & -(a+2b) & a & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & mb & -mb \end{pmatrix}.$$

The equations (5.19) become
$$-a\,v_0 + b\,v_1 = 0$$
$$a v_{j-1} - (a+jb)v_j + (j+1)b v_{j+1} = 0,\ j = 1, 2, \dots, m-1$$
$$a v_{m-1} - bm v_m = 0.$$

The solution of these equations can be obtained recursively. From the first equation
$$v_1 = (a/b)\,v_0.$$

Writing the second equation with $j = 1$ and putting there this value of v_1, we get
$$v_2 = \tfrac{1}{2}\,(a/b)^2\,v_0.$$

Proceeding in this way, we get
$$v_j = (1/j!)\,(a/b)^j\,v_0, j = 0, 1, 2, \dots, m.$$

Using $\sum_{j=0}^{m} v_j = 1$, one gets $v_0 = 1/(\sum_{i=0}^{m} (1/i!)\,(a/b)^i)$.

Formulas giving v_j are called ***Erlang's formulas*** (see Sec. 10.4D).

Example 5(e). *Machine Interference Problem.* Here there are m machines. Each of the machines operates independently and is serviced by a single servicing unit in case of break down. The operating time and servicing time of each machine are independently distributed as exponential distribution with parameters b and a respectively. Then the number of machines in operating condition at time t constitutes a Markov process $\{X(t), t \geqslant 0\}$ with state space $0, 1, \dots, m$.
(Incidentally, the rate matrix here is exactly the same as that of the preceding example; the limiting probabilities v_j are given by the Erlang's loss formula.)

Example 5(f). Two-channel service system. Consider Example 5(d) with $m = 2$. Suppose that the service channels are numbered I and II and suppose that we are interested in whether particular channels are busy (B) or free (F). Let the ordered pair (i, j) denote the state of the system, where i refers to that of the first channel and j to that of the second. The four states of the system (F, F), (F, B), (B, F) and (B, B) may be denoted by 0, 1, 2, 3 respectively. The process $\{X(t)\}$ denoting that the states of the system in terms of the two channels may be described by a Markov process with state space 0, 1, 2, 3.

Assume that when both the channels are free a demand may join either of the channels for service with equal probability ($\frac{1}{2}$). Thus, when both the channels are free, demand to each of the channels flow in accordance with a Poisson process with parameter $a/2$. We have

$$\begin{aligned}
p_{0,j}(\Delta t) &= (a/2)\,\Delta t + o\,(\Delta t), & j &= 1, 2\\
p_{1,j}(\Delta t) &= b\Delta t + o\,(\Delta t), & j &= 0\\
&= a\Delta t + o\,(\Delta t), & j &= 3\\
p_{2,j}(\Delta t) &= b\Delta t + o\,(\Delta t), & j &= 0\\
&= a\Delta t + o\,(\Delta t), & j &= 3\\
p_{3,j}(\Delta t) &= b\Delta t + o\,(\Delta t), & j &= 1, 2
\end{aligned}$$

and for all other combinations of $i \neq j$, $p_{i,j}(\Delta t) = o\,(\Delta t)$.

Thus the rate matrix A will be

$$\begin{pmatrix}
-a & a/2 & a/2 & 0\\
b & -(a+b) & 0 & a\\
b & 0 & -(a+b) & a\\
0 & b & b & -2b
\end{pmatrix}.$$

The normal equations (5.19) become

$$\begin{aligned}
-av_0 + bv_1 + bv_2 &= 0\\
(a/2)\,v_0 - (a+b)\,v_1 + bv_3 &= 0\\
(a/2)\,v_0 - (a+b)\,v_2 + bv_3 &= 0\\
av_1 + av_2 - 2bv_3 &= 0.
\end{aligned}$$

From the second and third equations we get $v_1 = v_2$ and then from the first we get $v_0 = (2b/a)\,v_1$ and from the last $v_3 = (a/b)\,v_1$. Utilising $\sum_{i=0}^{3} v_i = 1$, we at once get

$$v_1 = v_2 = \frac{ab}{a^2 + 2ab + 2b^2}$$

$$v_0 = \frac{2b^2}{a^2 + 2ab + 2b^2}$$

$$v_3 = \frac{a^2}{a^2 + 2ab + 2b^2}.$$

D. Graph Theoretic Approach for Determining V

The graph theoretic approach for determining V in case of discrete parameter Markov chain (see Sec. 3.5) is also applicable to finite, irreducible Markov processes with continuous parameter. In case of discrete Markov chains, arc "weights" correspond to the transition probabilities p_{ij}; in case of continuous Markov processes, the arc weights correspond to the transition densities $a_{ij} = (d/dt)\, p_{ij}(t)|_{t=0}$. The same result (Theorem 3.7) leading to steady-state solutions can be applied to finite-state discrete parameter Markov chains as well as to finite-state continuous parameter Markov processes, provided these are irreducible (Solberg). This is a definite advantage of this graph theoretic approach.

The procedure for continuous parameter process is illustrated below.

Example 5(g). *Erlang's formula*: Consider the machine interference (or m-channel loss system) problem discussed above.

Here for $i = 0, 1, 2, \ldots, m-1, j \neq i$

$$a_{i,j} = a, \qquad j = i + 1$$

and for $i = 1, 2, \ldots, m,$ $\qquad j \neq 1$

$$a_{i,j} = ib, \qquad j = i - 1$$

and

$$\begin{aligned} a_{i,i} &= -a, & i &= 0 \\ &= -mb, & i &= m \\ &= -(a + ib), & i &\neq 0, m. \end{aligned}$$

Also

$$a_{ij} = 0, \; j \neq i - 1, \; i \text{ or } i + 1.$$

The transition diagram may be shown as indicated below:

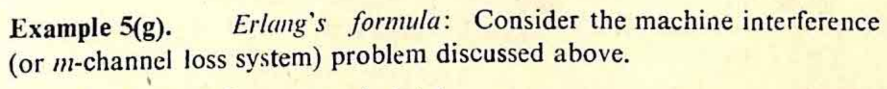


Fig. 4.3

Denote c_r = sum of weights of all the intrees to the point

$$r = 0, 1, 2, \ldots, m.$$

Here there is only one intree to each point. We have

$$c_0 = (mb)\,((m-1)\,b)\ldots(2b)\,(b) = m!\, b^m$$

$$c_1 = (mb)\,((m-1)\,b)\ldots(2b)\,a \;= m!\, ab^{m-1}/1!$$

$$\ldots \qquad \ldots \qquad \ldots$$

$$c_r = (mb)\,((m-1)\,b)\ldots((r+1)\,b)\,a^r = m!\, a^r b^{m-r}/r!,$$

so that $\sum_{r=0}^{m} c_r = m! \sum_{r=0}^{m} (a^r b^{m-r})/r!$. Thus

$$v_k = \frac{1}{k!}\left(\frac{a}{b}\right)^k \Bigg/ \sum_{r=0}^{m} \frac{1}{r!}\left(\frac{a}{b}\right)^r, \qquad k = 0, 1, 2, \ldots, m.$$

4.6 Erlang Process

A. Introduction

Consider a series of events E' which occur in accordance with a Poisson process $N(t)$ with mean $\lambda' t$. The random variable X_i which represents the interval between two successive occurrences has density $h(x) = \lambda' e^{-\lambda' x}$ having mean $1/\lambda'$. The interval separating k occurrences, i.e. the interval between the events numbered r and $r+k$ (for $k \geqslant 1$) is given by the sum $S_k = X_1 + \ldots + X_k$ of k independent exponential distributions each with mean $1/\lambda'$. S_k is a gamma variate (Theorem 1.10) with density function

$$\begin{aligned} f(t) &= \frac{(\lambda')^k\, t^{k-1}\, e^{-\lambda' t}}{\Gamma(k)}, && t > 0 \\ &= 0 && , t \leqslant 0. \end{aligned} \tag{6.1}$$

We have $E\{S_k\} = k/\lambda'$ and $\operatorname{var}\{S_k\} = k/(\lambda')^2$.

Suppose that every kth occurrence of E' is represented by a new series of events E, i.e. E occurs for the first time at the kth occurrence of E', E occurs for the second time at the $(2k)$th occurrence of E' and so on. If $X(t)$ is the number of occurrences of E in an interval of duration t, then we have

$$\begin{aligned} \Pr\{X(t) = n\} &= \Pr\{nk \leqslant N(t) \leqslant (n+1)\,k - 1\}, \quad n \geqslant 0 \\ &= \sum_{r=nk}^{(n+1)\,k-1} \Pr\{N(t) = r\} \\ &= \sum_{r=nk}^{(n+1)\,k-1} \frac{e^{-\lambda' t}\,(\lambda' t)^r}{r!}, \end{aligned} \tag{6.2}$$

since $N(t)$ is a Poisson process.

Evidently, for positive integral k,

$$\begin{aligned} \sum_{n=0}^{\infty} \Pr\{X(t) = n\} &= \sum_{n=0}^{\infty} \sum_{r=nk}^{(n+1)\,k-1} \frac{e^{-\lambda' t}\,(\lambda' t)^r}{r!} \\ &= \sum_{s=0}^{\infty} \frac{e^{-\lambda' t}\,(\lambda' t)^s}{s!} = 1. \end{aligned}$$

$\{X(t),\, t \geqslant 0\}$ is a renewal counting process with probability distribution given by (6.2). The interval between two successive occurrences of $X(t)$ which is the same as the interval between k occurrences of $N(t)$) is a gamma variate with density given by (6.1). For $k = 1$, $X(t)$ is identical with $N(t)$ and is a Poisson process.

(For the p.g.f. of $X(t)$ see Parzen p. 175.)

B. Erlangian Distribution

Putting $\lambda' = k\lambda$, we find that the interval separating k events of E' (which is the sum of k exponential variables each with mean $1/k\lambda$) has the density

$$\begin{aligned} f(t) &= \frac{(\lambda k)^k\, t^{k-1}\, e^{-k\lambda t}}{\Gamma(k)}, \quad t > 0 \qquad (6.3)\\ &= 0 \qquad\qquad\qquad\quad\; , t \leqslant 0 \end{aligned}$$

(obtained by putting $\lambda' = \lambda k$ in (5.1).)

The distribution having the above as its density function is called k-Erlangian or *Erlang-k* distribution (denoted by E_k), (see Example 3(f), Ch. 1).

If the occurrences of F are such that the intervals between two successive occurrences of F are identically and independently distributed with a common distribution E_k, then the number of occurrences $X(t)$, $(t \geqslant 0)$ of F in an interval of length t constitutes what is known as an *Erlang-k process*. $\{X(t)\}$ is a renewal process (Sec. 4.3F). We have

$$\Pr\{X(t) = n\} = \sum_{r=nk}^{(n+1)k-1} \frac{e^{-k\lambda t}\,(k\lambda t)^r}{r!}, \quad n \geqslant 0. \qquad (6.4)$$

(It may be said that a *random* selection preserves a Poisson process (property 3 of Poisson process Sec. 4.1) whereas a *deterministic* selection destroys it. A systematic selection of Poisson process results in an Erlang process.)

Example 6(a). Suppose that a system fails at the cumulative effect of n shocks. If shocks occur in accordance with a Poisson process with parameter λ', then the lifetime T of the system is a random variable being the sum of n independent exponential variates each with mean $1/\lambda'$. Hence T is a gamma variate with its density given by (6.1).

Example 6(b). Suppose that customers arrive at the desk of a receptionist in accordance with a Poisson process with a mean rate 2 (per minute). The receptionist immediately directs them alternately to two counters A and B so that each counter gets every second customer.

The interval T between two successive arrivals at A (or at B) is the sum of two independent exponential variates each with mean $\frac{1}{2}$ (minute). The density function of T is obtained by putting $k = 2$ and $\lambda' = 2$ in (6.1). We have

$$\begin{aligned} f(t) &= \frac{(2)^2\, t e^{-2t}}{\Gamma(2)}, \quad t > 0\\ &= 0, \qquad\qquad t \leqslant 0. \end{aligned}$$

$$E(T) = 2/2 = 1 \text{ and } \operatorname{var}(T) = 2/2^2 = 0.5$$

and $$\Pr(T > 1 \text{ minute}) = \int_1^\infty f(t)\,dt = 3e^{-2} = 0.406.$$

The probability $\Pr\{X(1) = 0\}$ that there is no arrival at B during the first minute is obtained by putting $t = 1$, $\lambda' = 2$, $k = 2$ and $n = 0$ in (6.2), i.e.

$$\Pr\{X(1)=0\} = \sum_{r=0}^{1} \frac{e^{-2}(2)^r}{r!} = 0.406.$$

In fact, $\Pr(T > 1) = \Pr\{X(1) = 0\}$.

The probability $\Pr\{X(1) = 1\}$ of 1 arrival is obtained by putting $n = 1$, i.e.

$$\Pr\{X(1)=1\} = \sum_{r=2}^{3} \frac{e^{-2}(2)^r}{r!} = 0.450.$$

Example 6(c). Suppose that particles reach a nuclear particle counter in accordance with a Poisson process with a mean rate of 4 (per minute), but that the counter records only every *third* particle actually arriving.

The probability that the number $X(t)$ of particles recorded in time t is n is given by (putting $k = 3$, $\lambda' = 4$ in (6.2))

$$\Pr\{X(t)=n\} = \sum_{r=3n}^{3(n+1)-1} \frac{e^{-4t}(4t)^r}{r!}, \quad n \geqslant 0.$$

If $t = 1, n = 1$, then

$$\Pr\{X(1)=1\} = \sum_{r=3}^{5} \frac{e^{-4}\,4^r}{r!} = 0.547.$$

The interval T between two successive records (i.e. recorded particles) has the density function (putting $\lambda' = 4$, $k = 3$ in (6.1)),

$$\begin{aligned} f(t) &= \frac{4^3 t^2 e^{-4t}}{\Gamma(3)}, \quad t > 0 \\ &= 0, \qquad\quad t \leqslant 0, \end{aligned}$$

with $E(T) = 3/4$ and $\operatorname{var}(T) = 3/16$.

The probability that the interval T between two records is less than or equal to $\frac{1}{2}$ (minute) is given by

$$\begin{aligned} \Pr(T \leqslant \tfrac{1}{2}) &= 1 - \Pr(T > \tfrac{1}{2}) \\ &= 1 - \Pr\{X(\tfrac{1}{2}) = 0\}, \\ &= 1 - \sum_{r=0}^{2} \frac{e^{-2}\,2^r}{r!} = 0.323. \end{aligned}$$

EXERCISES

4.1 Divide the interval $[0, t]$ into a large number n of small intervals of length h and suppose that in each small interval, Bernoulli trials with probability of success λh are held (i.e. trials with only two outcomes, success with probability λh and failure with probability $1 - \lambda h$). Show that the number of successes in an interval of length t is a Poisson process with mean λt. State the assumptions you make.

4.2 If $N_1(t)$, $N_2(t)$ are two independent Poisson processes with parameters λ_1, λ_2 respectively, then show that

$$\Pr\{N_1(t) = k \mid N_1(t) + N_2(t) = n\} = \binom{n}{k} p^k q^{n-k},$$

where $\quad p = \lambda_1/(\lambda_1 + \lambda_2), \quad q = \lambda_2/(\lambda_1 + \lambda_2).$

4.3 In a Poisson cluster process, suppose that each cluster X_i has the distribution of the geometric type:

$$p_k = \Pr\{X_i = k\} = q^{k-1}\, p, \quad k = 1, 2, \ldots.$$

Find the p.g.f., the mean and the variance of $M(t)$, (the total number of events). Show that the arrival stream of units is a renewal process with inter-arrival time distribution

$$A(t) = 1 - pe^{-\lambda t}\ (t \geqslant 0,\ 0 < p \leqslant 1).$$

4.4 The number of accidents in a town follows a Poisson process with a mean of 2 per day and the number X_i of people involved in the ith accident has the distribution

$$\Pr\{X_i = k\} = \frac{1}{2^k}\ (k \geqslant 1).$$

Find the mean and the variance of the number of people involved in accidents per week.

4.5 For the Pure birth process show that the interval T_k between the kth and $(k+1)$st birth has an exponential distribution with parameter λ_k.

4.6 *Zipf's rank size law*: Let Pr {a settlement S has n members at time t} $= p_n(t)$ and let Pr {a new member arrives in the settlement in $(t, t+h)$ given that there are n at time t} $= nah$, $a > 0$. Assume that the settlement starts with 1 individual at time 0, i.e. $p_1(0) = 1$ and that different types of individuals have different rates of arrival a, so that a is a random variable with density $g(a) = be^{-ab}$, $a > 0$, $b > 0$. Show that the probability distribution of S is given by

$$\Pr\{S = n \text{ at time } t\} = \int_0^\infty p_n(t)\, g(a)\, da = rB\,(n, r+1),\ n = 1, 2, \ldots,$$

where B is a Beta function and $r = b/t$. Show that for large n,

$$\Pr\{S \geqslant n \mid r\} \propto n^{-r}$$

(J.K. Ord (1975)).

4.7 Find the differential equation of pure death process. If the process starts with i individuals, find the mean and variance of the number $N(t)$ present at time t.

4.8 Show that for the linear growth process, the second moment $M_2(t)$ satisfies the differential equation:

$$M'_2(t) = 2(\lambda - \mu)\, M_2(t) + (\lambda + \mu)\, M(t).$$

Show that

$$\operatorname{var}\{X(t)\} = \{i(\lambda + \mu)/(\lambda - \mu)\}\, e^{(\lambda-\mu)t}\, \{e^{(\lambda-\mu)t} - 1\},\ \lambda \neq \mu$$
$$= 2i\mu t,\ \lambda = \mu.$$

i being the population size at $t = 0$.

4.9 Find the probability of ultimate extinction in the case of the linear growth process starting with i individuals at time 0.

4.10 Write down the differential-difference equations for the linear growth process with immigration

$$\lambda_n = n\lambda + a,\ \mu_n = n\mu.$$

Show that

$$M(t) = E\{X(t)\}$$

satisfies the differential equation

$$M'(t) = (\lambda - \mu)\, M(t) + a.$$

Hence show that with the initial condition $M(0) = i$, if $X(0) = i$, then

$$M(t) = \frac{a}{\lambda - \mu}\{e^{(\lambda-\mu)t} - 1\} + ie^{(\lambda-\mu)t},\ \lambda \neq \mu$$

and taking limit as $\lambda \to \mu$, show that

$$M(t) = at + i \text{ for } \lambda = \mu.$$

What is the limit of $M(t)$ as $t \to \infty$ for $\lambda < \mu$?

4.11 Continuation of 4.10: Linear growth process with immigration. Given the initial conditions $p_0(0) = 1$, $p_n(0) = 0$, $n \neq 0$, show that the population size, after the population has been developing for time T, has negative binomial distribution with index a/λ, and mean

$$\frac{a\{\exp(\lambda - \mu)\, T - 1\}}{\lambda - \mu}.$$

If $\lambda < \mu$ the population size will have a steady state distribution. When a is small compared to λ, show that, the population size, given that a population exists, follows a logarithmic series distribution.

[Williamson and Bretherton, *Ann. Math. Stat.* **35** (1964) 284–297.]

4.12 For the Yule-Fury process starting with only one individual at $t = 0$, show that the density function of T_n (the time at which the population *first* reaches the value $n\,(> 1)$ is given by

$$f(t) = \lambda\,(n - 1)\, e^{-\lambda t}\,(1 - e^{-\lambda t})^{n-2},\ t \geqslant 0.$$

Find $E(T_n)$. [Note that $T_n \leqslant t \Leftrightarrow N(t) \geqslant n$.]

4.13 *General form of Theorem 4.5:* For a Poisson process with parameter λ it is given that n occurrences take place in time t; show that the density function of the time of occurrence Z_k upto the kth event ($k < n$) is given by

$$f(x) = \frac{n!}{(k-1)!\,(n-k)!}\,\frac{x^{k-1}}{t^k}\left(1 - \frac{x}{t}\right)^{n-k},\quad 0 < x < t$$
$$= 0,\quad x \geqslant t.$$

Find the mean of Z_k.

4.14 In Example 6(*a*), let $n = 10$, $\lambda' = 1$ per hour; find the probability that the lifetime T (*i*) exceeds 10 hours, (*ii*) is less than or equal to 16 hours.

4.15 If $\{N(t), t \geqslant 0\}$ is a Poisson process, show that the process

$$X(t) = N(t + L) - N(t)$$

(where L is a positive constant) is covariance stationary.

4.16 Consider a Yule-Furry process starting with a single member at time $t = 0$, and having birth rate λ. Suppose that this first member (parent) is also subject to death, his lifetime being distributed as an exponential variable with parameter μ. Find the distribution of the number N of offspring due to this parent as well as his descendants at the time of death of the parent.

4.17 For a count process N of E' in E, where E and E' are independent Poisson processes with parameters λ and λ' respectively, show that the joint distribution of N and $T.T$ being the interval between two successive intervals of E, is given by

$$f(n, t) = \frac{\mu e^{-(\lambda+\mu)t} (\lambda t)^n}{n!}, \quad n=0, 1, 2, \ldots, t \geqslant 0.$$

Hence find the (marginal) distribution of N as

$$\Pr\{N = n\} = \int_0^\infty f(n, t)\, dt,$$

and the distribution of T as

$$f(t) = \sum_{n=0}^\infty f(n, t).$$

4.18 Suppose that $N_i(t)$, $i = 1, 2$ are two independent Poisson processes with parameters λ_i and $X(t) = N_1(t) + N_2(t)$ is their sum. Suppose that T_i and Z are the intervals *upto the first occurrences* of $N_i(t)$ and $X(t)$ respectively. Show that the random variable Z can be expressed in terms of T_i as $Z = \min(T_1, T_2)$. What is the disribution of Z?

How do you generalise the above for $i = 1, 2, \ldots, n$? Does the result hold if T_i, Z are the intervals between two successive occurrences of $N_i(t)$ and $X(t)$ respectively?

4.19 Write down the rate matrix for the birth and death process (Sec. 4.4). Assume that the limits $\lim_{t\to\infty} p_{ij}(t) = v_j$ exist.

Obtain the difference equations involving v_j from the Forward Kolmogorov equations. Using the method of induction, show that the solutions of these equations can be written as

$$v_{j+1} = \alpha_{j+1}\, v_0,$$

where $\quad \alpha_0 = 1$ and $\alpha_j = \prod_{r=1}^{j} \frac{\lambda_{r-1}}{\mu_r}, \; j = 1, 2, \ldots$

When does $\{v_j\}$ define a probability distribution?

4.20 Three-state process. Consider Example 5(*d*) with $m=2$. Find $p_{ij}(t)$. Obtain V from normal equation and also by using the graph theoretic technique.

4.21 Obtain V for Example 5(f) by using the graph theoretic technique.

4.22 Suppose that two machines working independently are looked after by one repairman. Suppose further that the break-downs of each of the machines occur in accordance with a Poisson process with parameter a, while the repair times of the two machines have independent exponential distributions with parameter b_1 and b_2 respectively. Suppose that $W(R)$ represents the working (under repair) state of a machine and the ordered pair (i, j) $(i. j = W, R)$ represents the state of the whole system so that the four states of the system are 0: (W, W), 1: (W, R), 2: (R, W) and 3: (R, R). Determine the rate matrix. Evaluate V

(i) from normal equations,

(ii) by applying the graph theoretic technique.

4.23 A repairable three-state process: The working of an automatic machine has been cited as an example of a three-state model by Proctor and Singh (1976). The states of an automatic machine on an assembly line could be good (state 0), (failed in mode 1 (state 1) and failed in mode 2 (state 2), (failure in mode 1 and 2 occurring due to carrying out some unwanted operations and due to ceasing to operate, respectively). If a_i is the constant failure rate to failure mode i and b_i the constant repair rate of failure mode i, then show that the rate matrix is

$$A = \begin{pmatrix} -(a_1+a_2) & a_1 & a_2 \\ b_1 & -b_1 & 0 \\ b_2 & 0 & -b_2 \end{pmatrix}.$$

If $p_i(t)$ is the probability that the device is in state i at time t, $i = 0, 1, 2$, then show that the differential-difference equations of the system are

$$p_0'(t) = -(a_1+a_2)\,p_0(t) + b_1 p_1(t) + b_2\, p_2(t)$$

$$p_i'(t) = a_i p_0(t) - b_i p_i(t),\ i = 1, 2.$$

If, at $t = 0$, the machine is in good state, i.e. $p_0(0) = 1$, $p_i(0) = 0$, $i = 1, 2$, then show that the solutions of the equations are (by the classical integration method or by applying Fourier transforms)

$$p_0(t) = 1 - C_1 - C_2 - (A_1 + A_2)\, e^{-rt} + (B_1 + B_2)\, e^{-st}$$

$$p_i(t) = C_i + A_i\, e^{-rt} - B_i\, e^{-st},\ i = 1, 2,$$

where $C_i = \dfrac{a_i b_{3-i}}{rs}$, $A_i = \dfrac{a_i(b_{3-i} - r)}{r(r-s)}$, $B_i = \dfrac{a_i(b_{3-i} - s)}{r(r-s)}$

and r, s are the roots of

$$x^2 - x\,(a_1 + a_2 + b_1 + b_2) + (b_1 b_2 + a_1 b_2 + a_2 b_1) = 0.$$

Show that, in steady state

$$p_0 = \frac{b_1 b_2}{rs},\ p_1 = \frac{a_1 b_2}{rs},\ p_2 = \frac{a_2 b_1}{rs}.$$

REFERENCES

N.T.J. Bailey, *Elements of Stochastic Processes, with Applications to the Natural Sciences*, Wiley, New York (1964).

M.S. Bartlett, *Introduction to Stochastic Processes*, Cambridge University Press, New York (1955).

A.T. Bharucha-Reid, *Elements of the Theory of Markov Processes and Their Applications*, McGraw-Hill, New York (1960).

A. Blanc-Lapierre and R. Fortet, *Théorie des Fonctions Aléatoires*, Masson et Cie, Paris (1953).

H. Cramér, "On streams of random events", *Skand. Akt. Supp.*, **52**, 13–23 (1969).

D.R. Cox and H.D. Miller, *The Theory of Stochastic Processes*, Methuen, London (1965).

W. Feller, *Introduction to Probability Theory and its Applications*, Vol. I, 3rd ed., Wiley, New York (1968).

M. Girault, *Stochastic Processes*, Springer-Verlag, Berlin (1966).

J. Grandell, *Doubly Stochastic Poisson Processes* Lecture notes in Maths., No. 529, Springer-Verlag, Berlin (1976).

S. Karlin, *A First Course in Stochastic Processes*, Academic Press, New York (1966).

J.F.C. Kingman, "Poisson counts for random sequence of events," *Ann. Math. Statistics*, **34**, 1217–32 (1963).

K. Ord, "The size of human settlements," in *A Modern Course on Statistical Distributions in Scientific Work*, Vol. **2**, Reidel, Dordrecht, Holland, 141–150 (1975).

E. Parzen, *Stochastic Processes*, Holden-Day, San Francisco (1962).

N.U. Prabhu, *Stochastic Processes*, Macmillan, New York (1965).

C.L. Proctor and B. Singh, "A repairable 3-state device", *IEEE Transactions on Reliability*, Vol. R–25, No. 3, 210–211 (1976).

10(45-95/1982)

CHAPTER 5

MARKOV PROCESSES WITH CONTINUOUS STATE SPACE

5.1 Introduction : Brownian Motion

Poisson process is a process in continuous time with a discrete state space. Here in a small interval of time Δt, there is either no change of state or there is only one change, the probability of more than one change being of the order of Δt. In this chapter we shall consider Markov processes such that in a small interval of time there is a small change of state or displacement. In such a process, changes of state are occurring continually all the time and the state space is continuous. An example of a continuous parameter continuous state space Markov process is the *Brownian motion process.* The process derives its name from the English botanist Robert Brown, who first noticed in 1827 the erratic movement of particles suspended in a fluid under the random impact of neighbouring particles. Einstein was the first to put forward a satisfactory mathematical model for the physical phenomenon in 1905, while Wiener, in 1923, gave it a rigorous mathematical form. Considerable attention has since been given which has resulted in remarkable progress in the subject. It has now wide applications in such fields as economics, communication theory, management science etc. Because of the connection with the theory of diffusion, Markov processes with continuous state space are also known as *diffusion processes.* A particle under diffusion or undergoing Brownian motion is also known as a Brownian particle. We shall confine ourselves to one-dimentional Brownian motion, that is, to motion along a fixed axis.

At epoch t, let $X(t)$ be the displacement along a fixed axis of a particle undergoing Brownian motion with $X(0)=x_0$. Consider an interval (s, t) of time; let us regard this interval as the sum of a large number of small intervals. The total displacement $X(t)-X(s)$ in this interval can be regarded as the limit of the sum of random displacements over the small intervals of time; these displacements may be interpreted as being caused by random walk performed by the Brownian particle over small intervals of time. Suppose that the random displacements are independently distributed. Then it can be seen that the central-limit theorem applies, whence it follows that the total displacement $X(t)-X(s)$ is normally distributed. Further, suppose that the displacement $X(t)-X(s)$ depends on the length of the interval (s, t) and not on the time-point s and that $X(t)-X(s)$ has the same distribution as $X(t+h)-X(s+h)$ for any $h>0$.

It is to be noted that here both time and space variables are continuous. The equations of the process obtained by taking limits of both time and space variables will be partial differential equations in both time and space variables. These equations, called diffusion equations, will be discussed in Sec. 5.3. In Sec. 5.2 we develop Wiener process as the continuous limit of the simple random walk.

It may be noted that there are some measure-theoretic subtleties involved in the passage from the discrete to the continuous case. Their considerations will, however, be beyond the scope of this book.

We assume that the process $X(t)$ is Markovian. Let the cumulative transition probability be

$$P(x_0, s; x, t) = \Pr\{X(t) \leqslant x \mid X(s) = x_0\},\ s < t \tag{1.1}$$

and let the transition probability density p be given by

$$p(x_0, s; x, t)\,dx = \Pr\{x \leqslant X(t) < x + dx \mid X(s) = x_0\}. \tag{1.2}$$

For a homogeneous process the transition probability depends only on the length of the interval $(t-s)$ and then the transition probability may be denoted in terms of the three parameters, $x_0, x, t-s$. We denote

$$\Pr\{x \leqslant X(t+t_0) < x + dx \mid X(t_0) = x\} \text{ by } p(x_0, x; t)\,dx$$

for any t_0. The Chapman-Kolmogorov equation can be written (as the continuous state analogue of (7.1) of Ch. 3) as follows:

$$P(x_0, s; x, t) = \int d_z\, P(x_0, s; z, v)\, P(z, v; x, t).$$

In terms of transition probabilities $p(x_0, s; x, t)$, we have

$$p(x_0, s; x, t) = \int p(x_0, s; z, v)\, p(z, v; x, t)\, dz.$$

5.2 Wiener Process

Consider that a (Brownian) particle performs a random walk such that in a small interval of time of duration Δt, the displacement of the particle to the right or to the left is also of small magnitude Δx, the total displacement $X(t)$ of the particle in time t being x. Suppose that Z_i is the random variable which denotes the length of the ith step taken by the particle in a small interval of time Δt and that

$$\Pr\{Z_i = \Delta x\} = p \quad \text{and} \quad \Pr\{Z_i = -\Delta x\} = q,\ p + q = 1$$

$0 < p < 1$, where p is independent of x and t.

Suppose that the interval of length t is divided into n equal subintervals of length Δt and that the displacements $Z_i, i = 1, \ldots, n$ in the n steps are mutually independent random variables. Then $n(\Delta t) = t$ and the total displacement $X(t)$ is the sum of n i.i.d. random variables Z_i, i.e.

$$X(t) = \sum_{i=1}^{n} Z_i.$$

We have

$$E\{Z_i\} = (p - q)\,\Delta x \quad \text{and} \quad \text{var}\,(Z_i) = 4pq\,(\Delta x)^2.$$

Hence

$$E\{X(t)\} = n\,E(Z_i) = t\,(p - q)\,\Delta x/\Delta t, \tag{2.1}$$

and

$$\text{var}\,X(t) = n\,\text{var}\,(Z_i) = 4pqt\,(\Delta x)^2/\Delta t.$$

To get a meaningful result, as $\Delta x \to 0$, $\Delta t \to 0$, we must have

$$\frac{(\Delta x)^2}{\Delta t} \to \text{a limit}, \; (p - q) \to \text{a multiple of } (\Delta x). \tag{2.2}$$

We may suppose, in particular, that in an interval of length t, $X(t)$ has mean-value function equal to μt and variance function equal to $\sigma^2 t$. In other words, we suppose that as $\Delta x \to 0$, $\Delta t \to 0$, in such a way that (2.2) are satisfied, and per unit time

$$E\{X(t)\} \to \mu \quad \text{and} \quad \text{var}\,\{X(t)\} \to \sigma^2. \tag{2.3}$$

From (2.1) for $t = 1$ and (2.3) we have

$$\frac{(p - q)\,\Delta x}{\Delta t} \to \mu; \; \frac{4pq\,(\Delta x)^2}{\Delta t} \to \sigma^2. \tag{2.4}$$

The relations (2.2) and (2.4) will be satisfied when

$$\Delta x = \sigma\,(\Delta t)^{1/2}, \tag{2.5a}$$

$$p = \tfrac{1}{2}\,(1 + \mu\,(\Delta t)^{1/2}/\sigma),\; q = \tfrac{1}{2}\,(1 - \mu\,(\Delta t)^{1/2}/\sigma). \tag{2.5b}$$

Now since Z_i are i.i.d. random variables, the sum $\sum_{i=1}^{n} Z_i = X(t)$ for large n is asymptotically normal with mean μt and variance $\sigma^2 t$ (by virtue of the central limit theorem for equal components). Note that here also t represents the length of the interval of time during which the displacement, that takes place is equal to the increment $X(t) - X(0)$. We thus find that for $0 < s < t$, $X(t) - X(s)$ is normally distributed with mean $\mu\,(t - s)$ and variance $\sigma^2\,(t - s)$. Further, the increments $X(s) - X(0)$ and $X(t) - X(s)$ are mutually independent; this implies that $X(t)$ is a Markov process.

We may now define a Wiener process or a Brownian motion process as follows:

The stochastic process $\{X(t),\, t \geqslant 0\}$ is called a *Wiener process* (or a *Wiener-Einstein process* or a *Brownian motion process*) with drift μ and variance parameter σ^2, if:

(i) $X(t)$ has independent increments, i.e. for every pair of disjoint intervals of time (s, t) and (u, v), where $s \leqslant t \leqslant u \leqslant v$, the random variables $X(t) - X(s)$ and $X(v) - X(u)$ are independent.

(ii) Every increment $X(t) - X(s)$ is normally distributed with mean $\mu\,(t - s)$ and variance $\sigma^2\,(t - s)$.

Note that (i) implies that Wiener process is a Markov process with independent increments and (ii) implies that a Wiener process is Gaussian.

Since $X(t) - X(0)$ is normally distributed with mean μt and variance $\sigma^2 t$, the transition probability density function p of a Wiener process is given by

$$p(x_0, x; t)\,dx = \Pr\{x \leqslant X(t) < x + dx \mid X(0) = x_0\}$$
$$= \frac{1}{\sigma\sqrt{(2\pi t)}} \exp\left\{-\frac{(x - x_0 - \mu t)^2}{2\sigma^2 t}\right\} dx. \qquad (2.6)$$

A Wiener process $X(t)$ with $X(0) = 0$, $\mu = 0$, $\sigma = 1$ is called a standard Wiener process.

5.3 Differential Equations for a Wiener Process

Let $X(t)$ be a Wiener process. We can consider the displacement in such a process as being caused by the motion of a particle undergoing displacements of small magnitude in a small interval of time. Suppose that $(t - \Delta t, t)$ is an infinitesimal interval of length Δt and that the particle makes in this interval a shift equal to Δx with probability p or a shift equal to $-\Delta x$ with probability $q = 1 - p$. Suppose that p and q are independent of x and t. Let the transition probability that the particle has a displacement from x to $x + \Delta x$ at epoch t, given that it started from x_0 at time 0, be $p(x_0, x; t)\,\Delta x$. Further suppose that $p(x_0, x; t)$ admits of an expansion in Taylor's series, i.e.

$$p(x_0, x \pm \Delta x; t - \Delta t) = p(x_0, x; t) - \Delta t\frac{\partial p}{\partial t} \pm \Delta x\frac{\partial p}{\partial x}$$
$$+ \tfrac{1}{2}(\pm\Delta x)^2\frac{\partial^2 p}{\partial x^2} + o(\Delta t). \qquad (3.1)$$

From simple probability arguments,we have

$$p(x_0, x; t)\,\Delta x = p.p(x_0, x - \Delta x; t - \Delta t)\,\Delta x$$
$$+ q.p(x_0, x + \Delta x; t - \Delta t)\,\Delta x. \qquad (3.2)$$

Making use of (3.1), and cancelling out the factor Δx from both sides of (3.2) we get

$$p(x_0, x; t) = p(x_0, x; t) - \Delta t\frac{\partial p}{\partial t} - \Delta x(p - q)\frac{\partial p}{\partial x}$$
$$+ \tfrac{1}{2}(\Delta x)^2\frac{\partial^2 p}{\partial x^2} + o(\Delta t).$$

Divide both sides by Δt. Using (2.4) and (2.5) and taking limits as $\Delta t \to 0$, $\Delta x \to 0$, we get

$$\frac{\partial}{\partial t}p(x_0, x; t) = -\mu\frac{\partial}{\partial x}p(x_0, x; t) + \tfrac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}p(x_0, x; t). \qquad (3.3)$$

This is a partial differential equation in both the variables x and t, being of first order in t and of the second order in x. The equation is known as the *forward diffusion equation* of the Wiener process. One can likewise obtain the *backward diffusion equation* of the process in the form

$$\frac{\partial}{\partial t}p(x_0, x; t) = \mu\frac{\partial}{\partial x_0}p(x_0, x; t) + \tfrac{1}{2}\sigma^2\frac{\partial^2}{\partial x_0^2}p(x_0, x; t). \qquad (3.4)$$

The solution of (3.3) as well as of (3.4) yields $p(x_0, x; t)$ as a normal density of the form given in (2.6) (see also Sec. 5.5). It may, however, be easily verified that $p(x_0, x; t)$ given by (2.6) satisfies (3.3) as well as (3.4). The equation for a Wiener process with drift $\mu = 0$ is known as the *heat equation.*

Note: The partial differential equation (3.3) [(3.4)] is known as the forward [backward] equation because it involves differentiation in $x[x_0]$. The reason why it is called diffusion equation is given in the next section.

It is to be noted that in sec. 5.2 and 5.3 we have made the following assumptions:

(i) in a small interval of time Δt, the displacement Δx is small (and that $\Delta x = 0\,((\Delta t)^{\frac{1}{2}})$;

(ii) $E\{X(t)\} \to \mu t$ in the limit;

(iii) $\text{var}\{X(t)\} \to \sigma^2 t$ in the limit.

The quantity μ in (ii) may also be interpreted as

$$\lim_{\Delta t \to 0} \frac{E\{X(t + \Delta t) - X(t)\}}{\Delta t} = \mu. \tag{3.5}$$

This implies that the infinitesimal mean (i.e. mean over Δt) of the expectation of the increment in $X(t)$ exists and is equal to a finite quantity μ.

The quantity σ^2 in (iii) may also be interpreted as

$$\lim_{\Delta t \to 0} \frac{E\{X(t + \Delta t) - X(t)\}^2}{\Delta t} = \sigma^2. \tag{3.6}$$

This implies that the infinitesimal mean (mean over Δt) of the variance of the increment in $X(t)$ exists and is equal to a finite quantity σ^2.

For a Wiener process, μ and σ^2 are assumed to be constants, independent of t or of x (where $X(t) = x$). By considering the transition mechanism with μ and σ^2 as functions of t or of x or both t and x, we get more general processes for which the equations corresponding to (3.3) and (3.4) will also be more general. We discuss below such equations.

5.4 Kolmogorov Equations

Let $\{X(t)\}$ be a Markov process in continuous time with continuous state space. We make the following assumptions: For any $\delta > 0$,

(i) $\Pr\{\,|X(t) - X(s)| > \delta \mid X(t) = x\} = o(t - s),\ s < t.$

In other words, small changes occur during small intervals of time.

(ii) $$\lim_{\Delta t \to 0} \frac{E\{X(t + \Delta t) - X(t) \mid X(t) = x\}}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \int_{|y-x| \leqslant \delta} (y - x)\, p(x, t; y, t + \Delta t)\, dy$$

$$= a(t, x).$$

In other words, the limit of the infinitesimal mean of the conditional expectation of the increment of $X(t)$ exists and is equal to $a(t, x)$, which is known as the *drift coefficient*.

$$\text{(iii)} \lim_{\Delta t \to 0} \frac{E\{[X(t+\Delta t) - X(t)]^2 \mid X(t) = x\}}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \int_{|y-x| \leqslant \delta} (y-x)^2 \, p(x, t; y, t+\Delta t)\, dy$$

$$= b(t, x).$$

In other words, the limit of the infinitesimal mean of the variance of the increment of $X(t)$ exists and is equal to $b(t, x)$, which is known as the *diffusion coefficient*.

A Markov process $X(t)$ satisfying the above conditions is known as a *diffusion process* and the partial differential equation satisfied by its transition p.d.f. is known as *diffusion equation*. We give below the equations (for derivation see Prabhu, Blanc-Lapierre and Fortet).

Let $X(t)$ be a Markov process satisfying (i), (ii) and (iii). If its transition p.d.f. $p(x_0, t_0; x, t)$ possesses continuous partial derivatives

$$\frac{\partial p}{\partial t}, \ \frac{\partial}{\partial x}(a(t, x)p), \ \frac{\partial^2}{\partial x^2}(b(t, s)p),$$

then $p(x_0, t_0; x, t)$ satisfies the forward Kolmogorov equation

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(a(t, x)p) + \tfrac{1}{2}\frac{\partial^2}{\partial x^2}(b(t, x)p). \tag{4.1}$$

This equation is also known as the *Fokker-Planck equation*. Suppose that $p(x_0, t_0; x, t)$ possesses continuous partial derivatives

$$\frac{\partial p}{\partial t_0}, \frac{\partial p}{\partial x_0}, \frac{\partial^2 p}{\partial {x_0}^2};$$

then $p(x_0, t; x, t)$ also satisfies the backward Kolmogorov equation

$$\frac{\partial p}{\partial t_0} = -a(t_0, x_0)\frac{\partial p}{\partial x_0} - \tfrac{1}{2} b(t_0, x_0)\frac{\partial^2 p}{\partial {x_0}^2}. \tag{4.2}$$

The diffusion equations for $X(t)$ were first derived by Kolmogorov. Feller showed that under suitable restrictions the equations admit of a unique solution. Fortet established some very interesting and important properties of the solutions.

Particular case: If the process is homogeneous, then

$$p(t_0, x_0; t, x) = p(x_0, x; t - t_0), \text{ and } a(t, x), b(t, x)$$

are independent of t.

If the process is additive, i.e. given that $X(t_0) = x_0$, the increment $X(t) - X(t_0)$, depends only on t_0 and t (and not on x_0), then

$$p(t_0, x_0; t, x) = p(x - x_0; t_0, t) \text{ and } a(t, x), b(t, x)$$

are independent of x.

The Kolmogorov equations, in these cases, can be easily deduced from the general equations.

5.5 First Passage Time Distribution for Wiener Process

The possible realizations of a stochastic process are called *sample paths* or *trajectories*. The structure and the properties of the sample paths of a Brownian motion or Wiener process are the subject matter of deep study. Without entering into the subtleties (which are beyond the scope of this work), we discuss here some results of Wiener process, using the property that the sample paths are continuous functions. We also make use of the simple but powerful 'reflection principle' (enunciated by D. André in 1887). The principle relates to the fact that there is a one-to-one correspondence between all paths from $A(a_1, a_2)$ to $B(b_1, b_2)$ which touch or cross the x-axis and all paths from $A'(a_1, -a_2)$ to B (see, Feller I for details). We shall first consider the following from which the distribution of the first passage time will be derived.

Distribution of the maximum of the Wiener process

Lemma: Let $X(t), 0 \leqslant t \leqslant T$ be a Wiener process with $X(0) = 0$ and $\mu = 0$. Let $M(T)$ be the maximum of $X(t)$ in $0 \leqslant t \leqslant T$, i.e. $M(T) = \max_{0 \leq t \leq T} X(t)$. Then for any $a > 0$

$$\Pr\{M(T) \geqslant a\} = 2 \Pr\{X(T) \geqslant a\}.$$

(This result was first obtained by Bachelier (1900).)

Proof: Consider the collection of sample paths $X(t), 0 \leqslant t \leqslant T$ such that $X(T) \geqslant a$. Now $X(0) = 0$ and $X(t)$ is continuous, there exists a time τ at which $X(t)$ first attains the value a (or $X(t)$ hits first the value a). The time T_a is itself a random variable. For $t > T_a$, $X_a(t)$ given below

$$X_a(t) = \begin{cases} X(t), & t < T_a \\ 2a - X(t), & t > T_a \end{cases}$$

gives the reflection of $X(t)$ about the line $x = a$. Note that $X_a(T) \leqslant a$, and that $M(T) \equiv \max_{0 \leq t \leq T} X(t) \geqslant a$ and $M_a(T) \equiv \max_{0 \leq t \leq T} X_a(t) \geqslant a$; further, by symmetry the sample paths $X(t)$ and $X_a(t)$ have the same probability of occurrence. From reflection principle, it follows that corresponding to every sample path $X(t)$ for which $X(T) \geqslant a$, there exist two sample paths such that $M(T) \geqslant a$. Further, its converse is also true, viz., every sample path $X(t)$ for which $M(T) \geqslant a$ corresponds to two sample paths $X(t)$ with equal probability, one of the paths being such that $X(T) > a$, unless $\{X(T) = a\}$, whose probability is zero. In fact, the set $\{M(T) \geqslant a\}$ is the union of three disjoint sets

$$\{M(T) \geq a,\ X(t) > a\},$$

$$\{M(T) \geq a,\ X(t) < a\},$$

$$\{M(T) \geq a,\ X(t) = a\}.$$

The probability of the third set is zero, while the other two are mapped onto one another by reflection about the line $x = a$ after the time T_a. Thus we have

$$\Pr\{M(T) \geq a\} = 2\Pr\{X(T) \geq a\}.$$

The above gives a heuristic proof of the lemma; as already indicated, a rigorous proof involves considerations beyond the scope of this book (see Karlin; Iosifescu and Tautu).

Let $X(t)$ be a Wiener process with $X(0) = 0$, $\mu = 0$ and let $M(t) = \max\limits_{0 \leq s \leq t} X(s)$. Then from the lemma, we get, for $t > 0$, $\sigma = 1$,

$$\Pr\{M(t) \geq a\} = 2\Pr\{X(t) \geq a\}$$

$$= \frac{2}{\sqrt{(2\pi t)}} \int_a^\infty \exp(-x^2/2t)\, dx \tag{5.1a}$$

$$= \frac{2}{\sqrt{(2\pi)}} \int_{a/\sqrt{t}}^\infty \exp(-y^2/2)\, dy \tag{5.1b}$$

(by changing the variable to $y = x/\sqrt{t}$)

$$= 2\left\{1 - \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{a/\sqrt{t}} \exp(-y^2/2)\, dy\right\}$$

$$= 2\{1 - \Phi(a/\sqrt{t})\}, \tag{5.1c}$$

Φ being the distribution function of the standard normal variate.

By changing the variable to $s = a^2t/x^2$, (5.1a) can be written as

$$\Pr\{M(t) \geq a\} = \frac{a}{\sqrt{(2\pi)}} \int_0^t s^{-3/2} \exp\left(-\frac{a^2}{2s}\right) ds,\ s > 0. \tag{5.1d}$$

Distribution of the first passage time to a fixed point

We can use the lemma to obtain the distribution of the random variable T_a, the first passage time to a fixed point a (> 0) (or the time of hitting a fixed point a first), for a Wiener process $X(t)$ with $X(0) = 0$, $\mu = 0$. The time T_a for $X(t)$ to hit the level a first will be less than t iff $M(t) = \max\limits_{0 \leq s \leq t} X(s)$ in that time is at least a. Thus for $t > 0$

$$\Pr\{M(t) \geq a\} = \Pr\{T_a \leq t\}. \tag{5.2}$$

$$F(t)=\frac{2}{\sqrt{(2\pi t)}}\int_a^\infty \exp(-x^2/2t)\,dx \tag{5.3a}$$

$$=2\,\{1-\Phi\,(a/\sqrt{t})\} \tag{5.3b}$$

$$=\frac{a}{\sqrt{(2\pi)}}\int_0^t s^{-3/2}\exp\left(-\frac{a^2}{2s}\right)ds,\ s>0. \tag{5.3c}$$

The density function of T_a is obtained by differentiating (5.3) with respect to t. Differentiation of (5.3c) (and also 5.3b) readily gives

$$f_{T_a}(t)=F'(t)=\frac{a}{\sqrt{(2\pi)}}\,t^{-3/2}\exp\left(-\frac{a^2}{2t}\right),\ t>0. \tag{5.4}$$

It may be easily verified that the Laplace transform $f^*(s)$ is

$$f^*(s)=\int_0^\infty e^{-st}f(t)\,dt$$

$$=\exp\{-a\sqrt{(2s)}\}\qquad \text{(see Exercise 1.10 (viii)).}$$

It can be seen that no moment of T_a exists finitely.

Let us find the density function of $M(t)$. The distribution function is

$$G(a)=\Pr\{M(t)\leqslant a\}$$

$$=1-\Pr\{M(t)>a\}$$

$$=1-2\Pr\{X(t)\geqslant a\}$$

$$=1-\frac{2}{\sqrt{(2\pi t)}}\int_a^\infty \exp(-x^2/2t)\,dx \tag{5.5a}$$

$$=1-2\,\{1-\Phi\,(a/\sqrt{t})\} \tag{5.5b}$$

$$=2\Phi\,(a/\sqrt{t})-1.$$

Differentiating (5.5a) or (5.5b) with respect to a, we get the density function

$$g_M(a)=G'(a)=\frac{2}{\sqrt{(2\pi t)}}\exp\left(-\frac{a^2}{2t}\right),\ a>0. \tag{5.6}$$

The results given above (for $\sigma=1$) can be suitably modified for any $\sigma>0$.

Note 1: We have obtained the distribution of T_a by using the lemma which gives a relation between the distributions of the Wiener process $X(t)$ and its maximum $M(t)$. However, the distribution of T_a can be obtained directly without bringing in the distribution of the maximum. In fact, it can be directly shown that, if $X(0)=0$, $a>0$, then

$$\Pr\{T_a\leqslant t\}=2\Pr\{X(t)\geqslant a\}. \tag{5.7}$$

For a proof of the above, see Prohorov and Rozanov (p. 245), who use conditional expectations to obtain the above. The distribution of the maximum $M(t)$ can then be obtained by using (5.2).

Note 2: For an alternative approach to the distribution of T_a using differential equations, see Cox and Miller (pp. 210, 220, 230), who obtain the distribution for any $\mu \geqslant 0$ and $\sigma > 0$. The density function $f_{T_a}(x)$ (for $X(0) = 0, \mu \geqslant 0, \sigma > 0$) of T_a is found to be

$$f_{T_a}(x) = \frac{a}{\sigma\sqrt{(2\pi x^3)}} \exp\left\{-\frac{(a-\mu x)^2}{2\sigma^2 x}\right\},\ x > 0; \tag{5.8}$$

and its Laplace transform is

$$f^*(s) = \exp\left[(a/\sigma^2)\left\{+\mu - \sqrt{(\mu^2 + 2s\sigma^2)}\right\}\right].$$

The mean and the variance for T_a for $\mu \neq 0$ are given by

$$E\{T_a\} = a/\mu \text{ and } \operatorname{var}\{T_a\} = a\,\sigma^2/\mu^3.$$

Note 3: The function (5.8) with $\mu > 0$ is the density function of the distribution of the first passage time of Brownian motion with a positive drift. This distribution having density function (5.8) is known as *inverse Gaussian distribution* because of the inverse relationship between the cumulant generating function of this distribution and that of normal distribution. Such a distribution was also obtained by Wald as the limiting form for the distribution of the sample size in a sequential probability ratio test. For properties of this distribution, see Johnson and Kotz (1970), and for statistical applications, see Folks and Chhikara (1978).

Example 5(a). Suppose that $\{X(t),\ 0 < t\}$, is a Wiener process with $X(0) = 0$, and $\mu = 0$. Then

$$\Pr\{X(t) \leqslant x\} = \Pr\{X(t)/\sigma\sqrt{t} \leqslant x/\sigma\sqrt{t}\} = \Phi(x/\sigma\sqrt{t}).$$

Consider the process

$$Y(t) = t\,X(1/t) \text{ in } 0 < t \leqslant 1 \text{ with } Y(0) = 0.$$

We have

$$E\{Y(t)\} = 0 \text{ and } \operatorname{var}\{Y(t)\} = t^2\,(\sigma^2/t) = \sigma^2 t.$$

Further,

$$\begin{aligned}\Pr\{Y(t) \leqslant y\} = \Pr\{t\,X(1/t) \leqslant y\} &= \Pr\left\{\frac{X(1/t)}{\sigma\sqrt{(1/t)}} \leqslant \frac{y/t}{\sigma\sqrt{(1/t)}}\right\} \\ &= \Phi\,(y/\sigma\sqrt{t}).\end{aligned}$$

Thus $Y(t)$, $0 < t \leqslant 1$ with $Y(0) = 0$ is also a Wiener process with $\mu = 0$ and variance $\sigma^2 t$.

Example 5(b). Consider a Wiener process $X(t)$ with $X(0) = 0$. Its mean value function is μt and variance function $\sigma^2 t$. For $0 < s < t$, the covariance function is

$$\begin{aligned}C(s,t) &= \operatorname{cov}\{X(s), X(t)\} = \operatorname{cov}\{X(s), X(s) + X(t) - X(s)\} \\ &= \operatorname{cov}\{X(s), X(s)\} + \operatorname{cov}\{X(s), X(t) - X(s)\} \\ &= \operatorname{cov}\{X(s), X(s)\},\end{aligned}$$

since the process has independent increments. Thus

$$C(s, t) = \text{var}\ \{X(s)\} = \sigma^2 s$$

(see also Exercise 2.7). The process is not convariance stationary even when $\mu = 0$.

Example 5(c). Suppose that $X(t), t > 0$ is a Wiener process with $X(0) = 0$. Its first passage time, T_a, to a has the same distribution as $1/u^2$, where u is a normal variate with mean 0 and s.d. σ/a. For, the distribution function of $1/u^2$ for $t > 0$, is

$$\begin{aligned} F(t) &= \Pr\{1/u^2 \leqslant t\} = \Pr\{u^2/(\sigma/a)^2 \geqslant a^2/\sigma^2\, t\} \\ &= \Pr\{(au/\sigma) \geqslant (a/\sigma\sqrt{t})\} + \Pr\{(au/\sigma) \leqslant -(a/\sigma\ \sqrt{t})\} \\ &= 1 - \Phi(a/\sigma\ \sqrt{t}) + \Phi(-a/\sigma\ \sqrt{t}) \\ &= 2\,(1 - \Phi(a/\sigma\ \sqrt{t})), \end{aligned}$$

which is the distribution function of T_a (see equation (5.3b)). The distribution of $1/u^2$ is *inverse Gaussian.*

Example 5(d). Let $X(t)$ be a Wiener process with $\mu = 0$ and $X(0) = 0$. To find the distribution of T_{a+b} for $0 < a < a + b$.

Suppose t_a is a value of T_a, i.e. $X(t)$ reaches the level a for the first time at epoch t_a. We may then consider that the process starts at (t_a, a) and reaches the level $(a + b)$, which is b units higher than a. Suppose that $T_{a+b} - T_a$ is the duration of the interval at the ends of which $X(t)$ first reaches the level a and then reaches first the level $a + b$. Then T_a and $(T_{a+b} - T_a)$ are independent random variables denoting first passage times to a and b respectively. The L.T. of the p.d.f. of T_a is

$$h_a{}^*(s) = \exp\{-a\sqrt{(2s)}/\sigma\}$$

and that of T_b is

$$h_b{}^*(s) = \exp\{-b\ \sqrt{(2s)}/\sigma\}.$$

Thus the L.T. of the p.d.f. of T_{a+b} is

$$h_a{}^*(s)\ h_b{}^*(s) = \exp\{-(a+b)\ \sqrt{(2s)}/\sigma\} = h^*_{a+b}(s).$$

The distribution of T_{a+b} has the same form as that of T_a and can be obtained by replacing a by $(a + b)$ in the distribution of T_a.

5.6 Ornstein-Uhlenbeck Process

We have seen that for a Wiener process $X(t)$, the displacement Δx in a small interval of time Δt is also small, being of $0(\sqrt{(\Delta t)})$. The velocity which is of $0(\sqrt{(\Delta t)}/\Delta t) = 0(1/\sqrt{(\Delta t)})$ tends to infinity as $\Delta t \to 0$. Thus the Wiener process does not provide a satisfactory model for Brownian motion for small values of t, although for moderate and large values of t it does so. As alternative model which holds for small

t was proposed by Ornstein and Uhlenbeck in 1930. Here instead of the displacement $X(t)$, the velocity $U(t) = X'(t)$ at time t is considered.

The equation of motion of a Brownian particle can be written as

$$dU(t) = -\beta\, U(t)\, dt + dF(t), \tag{6.1}$$

where $-\beta U(t)$ represents the systematic part due to the resistance of the medium and $dF(t)$ represents the random component. It is assumed that these two parts are independent and that $F(t)$ is a Wiener process with drift $\mu = 0$ and variance parameter σ^2. The Markov process $U(t)$ is such that in a small interval of time the change in $U(t)$ is also small. Since $F(t)$ is a Wiener process, we have from (6.1)

$$\begin{aligned}
&\lim_{\Delta t\to 0} \frac{E\{U(t+\Delta t) - U(t) \mid U(t) = u\}}{\Delta t}\\
&= -\beta u + \lim_{\Delta t\to 0} \frac{E\{\Delta F(t)\}}{\Delta t}\\
&= -\beta u,
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{\Delta t\to 0} \frac{\operatorname{var}\{U(t+\Delta t) - U(t) \mid U(t) = u\}}{\Delta t}\\
&= \lim_{\Delta t\to 0} \frac{(\Delta t)^2}{\Delta t} + \lim_{\Delta t\to 0} \frac{\operatorname{var}\{\Delta F(t)\}}{\Delta t},\\
&= \sigma^2.
\end{aligned}$$

In other words, the limits exist. So the process $U(t)$ is a diffusion process and its transition p.d.f. $p(u_0; u, t)$ satisfies the forward Kolmogorov equation (4.1) with $a(u, t) = -\beta u$ and $b(u, t) = \sigma^2$. That is, p satisfies the differential equation

$$\frac{\partial p}{\partial t} = \beta \frac{\partial}{\partial u}(up) + \tfrac{1}{2}\sigma^2 \frac{\partial^2 p}{\partial u^2}. \tag{6.2}$$

Let us assume that $U(0) = u$ and that as $u_0 \to \pm\infty$, $p \to 0$ and $(\partial p/\partial u) \to 0$. The solution of (6.2) gives p, the transition p.d.f. of $U(t)$. It is more convenient to consider the equation corresponding to (6.2) in terms of the characteristic function of p, i.e.

$$\phi(u_0; \theta, t) = \int_{-\infty}^{\infty} e^{i\theta u}\, p(u_0; u, t)\, du.$$

We have

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{i\theta u} \frac{\partial}{\partial u}(up)\, du &= e^{i\theta u}\, up \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} i\theta\, e^{i\theta u}\, up\, du\\
&= -\theta \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} e^{i\theta u}\, p\, du\\
&= -\theta \frac{\partial \phi}{\partial \theta};
\end{aligned}$$

$$\int_{-\infty}^{\infty} e^{i\theta u}\, \frac{\partial^2 p}{\partial u^2} = e^{i\theta u}\, \frac{\partial p}{\partial u} \Bigg|_{-\infty}^{\infty} - i\theta \int_{-\infty}^{\infty} e^{i\theta u} \frac{\partial p}{\partial u}\, du$$

$$= - i\theta \left\{ e^{i\theta u}\, p \Bigg|_{-\infty}^{\infty} - i\theta \int_{-\infty}^{\infty} e^{i\theta u}\, p\, du \right\}$$

$$= - \theta^2 \phi.$$

The equation (6.2) then becomes

$$\frac{\partial \phi}{\partial t} + \beta\theta\, \frac{\partial \phi}{\partial \theta} = - \tfrac{1}{2}\sigma^2\theta^2\phi. \tag{6.3}$$

The equation (6.3) is of Lagrange type. It can be shown that

$$\phi(u_0;\, \theta,\, t) = \exp\,\{i\theta\, u_0\, e^{-\beta t} - \frac{1}{4\beta}\, \theta^2\sigma^2\, (1 - e^{-2\beta t})\}. \tag{6.4}$$

This is the characteristic function of normal distribution with $m(t) = u_0 e^{-\beta t}$ and variance function $\sigma^2(t) = \sigma^2\,(1 - e^{-2\beta t})/4\beta$.

In other words, the transition p.d.f. p is normal with mean value function $m(t)$ and variance function $\sigma^2(t)$ and p can be written as:

$$p(u_0;\, u,\, t) = \frac{1}{\sqrt{(2\pi\sigma^2(t))}}\, \exp\,\{ - (x - m(t))^2/2\sigma^2(t)\}. \tag{6.5}$$

Thus the process $U(t)$ is a Gaussian process with mean value function $m(t)$ and variance function $\sigma^2(t)$. $U(t)$ is a Markov process but it does not possess independent increments like the Wiener process. $U(t)$ is known as *Ornstein-Uhlenbeck process (O-U.P.)*. For large t, $m(t) \to 0$ and $\sigma^2(t) \to \sigma^2/2\beta$, i.e. the distribution of velocity is normal with mean 0 and variance $\sigma^2/2\beta$. We thus get an equilibrium distribution and $U(t)$ is said to in statistical equilibrium. For small t, $m(t) \to u_0$ and $\sigma^2(t) \to \sigma^2 t$.

Example 6(a). Joint distribution of $U(t)$ and $U(t + \tau)$ when $U(t)$ is in equilibrium. For large t, the limiting distribution of $U(t)$ is normal with mean 0 and variance $\sigma^2/2\beta = {\sigma_0}^2$ (say). The conditional distribution of $U\,(t + \tau)$, given $U\,(t) = u$, is normal with mean $ue^{-\beta\tau}$ and variance $\sigma^2(\tau) = \sigma^2(1 - e^{-2\beta\tau})/2\beta$. Thus the unconditional distribution of $U\,(t + \tau)$ has the following density

$$h(x) = \frac{1}{\sqrt{(2\pi{\sigma_0}^2)}} \int_{-\infty}^{\infty} \left[\exp\left(- \frac{{x_0}^2}{2{\sigma_0}^2} \right) \frac{1}{\sqrt{(2\pi\,\sigma^2(t))}} \right.$$

$$\left. .\exp\left\{ - \frac{1}{2\sigma^2(t)}\,(x - x_0\, e^{-\beta t})^2 \right\} \right] dx_0$$

$$= \frac{1}{\sqrt{(2\pi\,{\sigma_0}^2)}}\, \exp\left(- \frac{x^2}{2{\sigma_0}^2} \right).$$

Thus the unconditional distribution of $U(t + \tau)$ is Gaussian, has mean 0 and variance σ_0^2 and the unconditional distribution of $U(t + \tau)$ is the same as the equilibrium distribution of $U(t)$. The joint distribution of $U(t)$ and $U(t+\tau)$ has the density

$$f(x,\ y) = \frac{1}{\sqrt{(2\pi\ \sigma_0{}^2)}} \exp\left(-\frac{x^2}{2\sigma_0{}^2}\right) \cdot \frac{1}{\sqrt{(2\pi\ \sigma^2(t))}} \exp\left\{-\frac{1}{2\sigma^2(t)}(y - xe^{-\beta t})^2\right\}$$

$$= \frac{1}{2\pi\sigma_0{}^2\sqrt{(1-c^2)}} \exp\left\{-\frac{1}{2\sigma_0{}^2(1-c^2)}(y^2 - 2axy + x^2)\right\},$$

where $c = e^{-\beta t}$.

It follows that $U(t)$ and $U(t+\tau)$ have a bivariate Gaussian distribution with

$$E\{U(t)\} = E\{U(t+\tau)\} = 0,$$

$$\text{cov}\ \{U(t),\ U(t+\tau)\} = \sigma_0^2\ e^{-\beta|\tau|}$$

and
$$\text{var}\ \{U(t)\} = \sigma_0^2.$$

The mean and the variance of $U(t)$ are finite and the covariance function $\text{cov}\ \{U(t),\ U(t+\tau)\}$ is a function of the absolute difference only. Hence $U(t)$ is covariance stationary. Again, since $U(t)$ is Gaussian, $U(t)$ is strictly stationary.

Note that Wiener process is not covariance stationary.

Example 6(b). *The O-U.P. as a transformation of a Wiener process:* Let $X(t),\ t \geqslant 0$, be a standard Wiener process. Let

$$Y(t) = \frac{1}{\sqrt{g(t)}}\ X(a\ g(t)),\ \ a > 0,$$

and let the (non-random) function $g(t)$ be positive, strictly increasing with $g(0) = 1$.

We have
$$E\{Y(t)\} = 0$$

and
$$\text{var}\ \{Y(t)\} = \frac{1}{g(t)}\ \text{var}\ \{X(a\ g(t))\}$$

$$= \frac{1}{g(t)}\ \{a\ g(t)\} = a.$$

Since $(X(t_1), \ldots, X(t_n))$ is multivariate normal, so also is $(X(a\ g(t_1)), \ldots, X(a\ g(t_n))$, and for $\tau > 0$,

$$\text{cov}\{Y(t),\ Y(t+\tau)\} = \text{cov}\left\{\frac{X(a\ g(t)),\ X(a\ g(t+\tau))}{(g(t)\ g(t+\tau))^{\frac{1}{2}}}\right\}$$

$$= \text{var}\ \{X(a\ g(t))\}/\{(g(t)\ g(t+\tau))\}^{\frac{1}{2}},$$

since, for the Wiener process $X(t)$, $\text{cov}\ \{X(t),\ X(t+\tau)\} = \text{var}\ X(t)$ (see Example 5(b)).

Now,
$$\text{var}\ \{X(a\ g(t))\} = \text{var}\ \{\sqrt{g(t)}\ Y(t)\}$$

$$= g(t)\ \text{var}\ \{Y(t)\} = a\ g(t)$$

and thus
$$\text{cov}\ \{Y(t),\ Y(t+\tau)\} = a\ \{g(t)/g(t+\tau)\}^{\frac{1}{2}}$$

and for $t = 0$, the covariance equals $a/\{g(\tau)\}^{\frac{1}{2}}$.

The process $Y(t)$, which has finite mean and variance, will be covariance stationary provided $\operatorname{cov}\{Y(t), Y(t+\tau)\}$ depends only on τ. Thus we must have $g(t+\tau) = g(\tau).g(t)$, and in order to satisfy this equation, $g(t)$ must be an exponential function, say, $g(t) = e^{2\beta t}$ $(\beta > 0)$. We have then

$$\operatorname{cov}\{Y(t), Y(t+\tau)\} = a\, e^{-\beta\tau}, \ (\tau > 0),$$

and thus we find that for $a > 0, \beta > 0$

$$Y(t) = e^{-2\beta t}\, X(a\, e^{2\beta t})$$

is a stationary Gaussian Markov process. In other words, $Y(t)$ has the structure of an Ornstein-Uhlenbeck process.

EXERCISES

5.1 If $X(t)$, with $X(0)$ and $\mu = 0$, is a Wiener process, show that $Y(t) = \sigma X(t/\sigma^2)$ is also a Wiener process. Find its covariance function.

5.2 If $X(t)$ with $X(0) = 0$ and $\mu = 0$ is a Wiener process and $0 < s < t$, show that for at least one τ satisfying $s \leqslant \tau \leqslant t$,

$$\Pr\{X(\tau) = 0\} = (2/\pi)\cos^{-1}((s/t)^{1/2}).$$

5.3 Let $X(t)$, with $X(0) = 0$, be a standard Wiener process and let T_a be the first passage time of $X(t)$. Show that T_a and $a^2\, T_1$ are identically distributed. If Z_i, $i = 1, 2, \ldots, n$ are i.i.d. as T_1 then show that $\Sigma\, Z_i/n$ and Z_1 are identically distributed.

5.4 *Additive process*: Here

$$a(t, x) = a(t) \text{ and } b(t, x) = b(t).$$

Show that the Kolmogorov forward equation can be reduced to

$$\frac{\partial p}{\partial t_1} = \frac{1}{2}\frac{\partial^2 p}{\partial x_1^2}$$

by a suitable change of variables, and that the corresponding distribution is Gaussian.

5.5 Let $X(t)$ be the displacement process corresponding to the velocity process $U(t)$ (O-U. process). Show that in equilibrium,

$$E\{X(t) - X(0)\} = 0$$

and $\quad \operatorname{var}\{X(t) - X(0)\} = \sigma^2(\beta t - 1 + e^{-\beta t})/\beta^3.$

Deduce that,

$$\operatorname{var}\{X(t) - X(0)\} \sim \sigma^2 t^2/2\beta \text{ for small } t$$
$$\sim \sigma^2 t/\beta^2 \text{ for large } t.$$

5.6 Show that if $Y(t)$ is a $O\text{-}U.P.$ with mean value 0 and covariance function $C\{Y(s), Y(t)\} = ae^{-\beta|t-s|}$, then

$$X(t) = (a/t)^{-1/2}\, Y(\tfrac{1}{2}\beta \log (t/a))$$

is a standard Wiener process.

5.7 Suppose $U(t)$ is an $O\text{-}U.$ process with $E\{U(t)\} = 0$, $\operatorname{cov}\{U(s), U(t)\} = \sigma_0^2 e^{-\beta|t-s|}$ and $X(t)$, (where $(d/dt)\, X(t) = U(t)$) gives the position of the particle.

Show that the conditional distribution of $X(t)$ $(t > 0)$ given that $X(0) = x_0$ is normal with mean $E\{X(t)\} = x_0$ and $\operatorname{var}\{X(t)\} = (2\sigma_0^2/\beta^2)(e^{-\beta t} - 1 + \beta t)$ (Prabhu (1965)).

5.8 *Inverse Gaussian Process:* Consider a Wiener process $X(t)$ with positive drift. Without loss of generality, suppose that $X(0)=0$,, $\mu=1$, $\sigma = 1$. Then T_a, the first passage time to a given point a $(a > 0)$ has density

$$f(a, x)\, dx = \frac{a}{\sqrt{(2\pi x^3)}} \exp\{-(x-a)^2/2x\}\, dx, \quad x > 0.$$

Instead of a given point a, consider a variable point t (> 0). Then T_t has density

$$f(t, x)\, dx = \frac{t}{\sqrt{(2\pi x^3)}} \exp\{-(x-t)^2/2x\}\, dx.$$

This inverse Gaussian distribution has mean t and variance t. A new stochastic process having f for its distribution can now be defined.

An *inverse Gaussian process* (or a *first passage time distribution process*) is a stochastic process $\{X(t), t \geqslant 0\}$, with $X(0) = 0$ such that

(1) $X(t)$ has independent increments, i.e. for every pair of disjoint intervals of time (s, t) and (u, v), where $s \leqslant t \leqslant u \leqslant v$, the random variable $X(t) - X(s)$ and $X(v) - X(u)$ are independent.

(2) Every increment $X(t+h) - X(h)$ has inverse Gaussian distribution with mean t and variance t.

For $\quad s < t, \quad E\{X(t) - X(s)\} = (t-s)$,

$$\operatorname{var}\{X(t) - X(s)\} = (t-s)$$

and $\quad C(s, t) = \operatorname{cov}\{X(s), X(t)\} = \operatorname{var} X(s))$

(Wasan, 1969).

5.9 *Gamma Process*: If $Y_i(t)$, $i = 1, 2, \ldots, m$ are i.i.d. Gaussian processes with zero mean value function and covariance function

$$\operatorname{cov}\{Y_i(s), Y_i(t)\} \equiv C(s, t) \text{ for all } i,$$

then $X(t) = \frac{1}{2} \sum_{i=1}^{m} Y_i^2(t)$ is a Gamma process with parameter $m/2$.

Show that

$$E\{X(t)\} = \tfrac{1}{2} m\, C(t, t),$$

$$\operatorname{cov}\{X(t), X(s)\} = \tfrac{1}{2}m\,(C(t, t))^2$$

and that the distributions of a Gamma process are gamma distributions.

5.10 *Negative binomial process*: In our discussion on Poisson process we consider λ as a constant or as a non-random function $\lambda(t)$ of time. It is natural to introduce some kind of random variations of λ also; we may consider λ to be a random variable or as a random function (or stochastic process) $\lambda(t)$. A Poisson process whose intensity function $\lambda(t)$ is a stochastic process is called a conditional Poisson process or a doubly stochastic Poisson process. Several studies have been made on conditional and mixed Poisson process (see also Sec. 4.3E).

If the parameter λ of a Poisson distribution is a random variable having gamma distribution, then the mixed Poisson distribution is negative binomial. By analogy, a negative binomial process $N(t)$ is defined as a mixed Poisson process whose intensity function $\lambda(t)$ is a Gamma process. It may be remarked that the distributions of a negative binomial process are not necessarily negative bionmial.

Suppose $Y_i(t)$ are i.i.d. *O-U.* processes with $E(Y_i) = 0$ and covariance function $c(s, t) = \sigma^2 e^{-\beta|t-s|}$. Then $\frac{1}{2}\sum_{i=1}^{m} Y_i^2(t)$ is a Gamma process and the mixed Poisson process, whose intensity function $\lambda(t)$ is a Gamma process, is a negative binomial process $N(t)$.

Examine whether the distributions of $N(t)$ are negative bionomial.

Show that

$$E\{N(t)\} = \tfrac{1}{2}m\sigma^2 t$$

and $\operatorname{var}\{N(t)\} = \frac{1}{2}m\sigma^2 t + m\sigma^4\,(e^{-2\beta t} - 1 + 2\beta t)/4\beta^2$

(Barndorff-Nielsen and Yeo, 1969).

REFERENCES

O. Barndorff-Nielson and G.F. Yeo, "Negative binomial processes," *J. Appl. Prob.* 6, 633-647 (1969).

A. Blanc-Lapièrre and R. Fortet, *Théorie des Fonctions Aléatories,* Masson et Cie., Paris (1953).

D.R. Cox and H.D. Miller, *The Theory of Stochastic Processes*, Methuen, London (1965).

J.L. Doob, *Stochastic Processes,* Wiley, New York (1953).

A. Einstein, *Investigation on the Theory of Brownian Movement*, Dover, New York (1956).

J.L. Folks and R.S. Chhikara, "The inverse Gaussian distribution and its statistical application—a review," *J. R. S. S.*, **B. 40**, 263-275 (1978).

M. Iosifescu and P. Tautu, *Stochastic Processes and Applications in Biology and Medicine,* I, Springer-Verlag (1973).

N.L. Johnson and S. Kotz, *Continuous Univariate Distributions*, Vol. I. Houghton Mifflin Co., Boston (1970).

S. Karlin, *A First Course in Stochastic Processes*, Academic Press, New York (1966).

P. Lévy, *Processus Stochastiques et Mouvement Brownien*, Gauthier-Villars, Paris (1953).

N.U. Prabhu, *Stochastic Processes*, Macmillan, New York. (1965).

Yu. V. Prohorov and Yu. A. Rozanov, *Probability Theory*, Springer-Verlag (1969).

M.T. Wasan, "Sufficient conditions for a first passage time to be that of a Brownian motion," *J. Appl. Prob.* **6**, 218–223 (1969).

CHAPTER 6

RENEWAL PROCESSES AND THEORY

6.1 Renewal Processes

Renewal process as a generalisation of Poisson process was introduced in Sec. 4.3 F. Here we propose to discuss the topic in some details.

Let $\{X_n, n = 1, 2, \ldots\}$ be a sequence of non-negative independent random variables. Assume that $\Pr\{X_n = 0\} < 1$, and that the random variables are identically distributed with a distribution function $F(\cdot)$. Since X_n is non-negative, it follows that $E\{X_n\}$ exists and let us denote

$$E(X_n) = \int_0^\infty x \, dF(x) = \mu,$$

where μ may be infinite. Whenever $\mu = \infty$, $1/\mu$ shall be interpreted as 0.

Let $\quad S_0 = 0, S_n = X_1 + X_2 + \ldots + X_n; n \geqslant 1$

and let $F_n(x) = \Pr\{S_n \leqslant x\}$ be the distribution function of S_n.

We confine here to the more important case of continuous X_i.

Definition. Define the random variable

$$N(t) = \sup\{n : S_n \leqslant t\}. \tag{1.1}$$

The process $\{N(t), t \geqslant 0\}$ is called a *renewal process* with distribution F (or generated or induced by F).

It is also customary to say that the sequence of random variables $\{S_n, n = 1, 2, \ldots\}$ (as also the sequence $\{X_n, n = 1, 2, \ldots\}$) constitutes a renewal process with distribution F.

If, for some n, $S_n = t$, then a renewal is said to occur at t; S_n gives the time (epoch) of the nth renewal, and is called nth *renewal* (or *regeneration*) epoch. The random variable $N(t)$ gives the number of renewals occurring in $[0, t]$. The random variable X_n gives the interarrival time (or waiting time) between $(n-1)$th and nth renewals. The interarrival times are independently and identically distributed. When they have common exponential distribution, we get Poisson process as a particular case.

Simple Examples

One of the simplest examples of a renewal process is provided by lifetime distributions of a component such as an electric bulb which either works or fails completely. Suppose that the detection of the failure of a bulb and its replacement by a new bulb take place instantaneously and suppose that the lifetimes of bulbs are i.i.d. random variables with distribution F. We then have a renewal process with distribution F.

Consider a stage in an industrial process relating to production of a certain component in batches. Immediately on completion of production of a batch, that of another batch is undertaken. Suppose that the times taken to produce successive batches are i.i.d. random variables with distribution F. We get a renewal process with distribution F, where failure implies completion of production of a batch.

The function $M(t) = E\{N(t)\}$ is called the *renewal function* of the process with distribution F. It is clear that

$$N(t) \geqslant n \Leftrightarrow S_n \leqslant t \tag{1.2}$$

i.e.

$$N(t) \geqslant n \text{ if and only if } S_n \leqslant t.$$

Theorem 6.1. The distribution of $N(t)$ is given by

$$p_n(t) = \Pr\{N(t) = n\} = F_n(t) - F_{n+1}(t) \tag{1.3}$$

and the expected number of renewals by

$$M(t) = \sum_{n=1}^{\infty} F_n(t). \tag{1.4}$$

Proof: We have

$$\begin{aligned}\Pr\{N(t) = n\} &= \Pr\{N(t) \geqslant n\} - \Pr\{N(t) \geqslant n+1\}\\ &= \Pr\{S_n \leqslant t\} - \Pr\{S_{n+1} \leqslant t\}\\ &= F_n(t) - F_{n+1}(t).\end{aligned}$$

Again,

$$\begin{aligned}M(t) = E\{N(t)\} &= \sum_{n=0}^{\infty} n\, p_n(t)\\ &= \sum_{n=0}^{\infty} n\{F_n(t) - F_{n+1}(t)\}\\ &= \sum_{n=1}^{\infty} F_n(t)\\ &= \sum_{n=1}^{\infty} \Pr\{S_n \leqslant t\}. \text{ Hence proved. } \blacktriangle\end{aligned}$$

The relation (1.4) can be put in terms of Laplace transforms as follows:

Let $F'(x) = f(x)$ be the density function (p.d.f.) of X_n and $g^*(s)$ denote the Laplace transform of a function $g(t)$. Then taking Laplace transform of both sides of (1.4), we get

$$\begin{aligned}M^*(s) &= \sum_{n=1}^{\infty} F_n^*(s) = \frac{1}{s}\sum_{n=1}^{\infty} f_n^*(s)\\ &= \frac{1}{s}\sum_{n=1}^{\infty} [f^*(s)]^n = \frac{f^*(s)}{s\,[1 - f^*(s)]}.\end{aligned} \tag{1.5}$$

This is equivalent to

$$f^*(s) = \frac{s\, M^*(s)}{1 + s\, M^*(s)}. \tag{1.6}$$

These show that $M(t)$ and $F(x)$ can be determined *uniquely* one from the other.

Note: $M(t) = E\{N(t)\}$ is a *sure* function and *not* a random function or stochastic process.

Example 1(a). Let X_n have gamma distribution having density

$$f(x) = \frac{a^k\, x^{k-1}\, e^{-ax}}{(k-1)!}, \quad x \geqslant 0.$$

Then,

$$f^*(s) = \left(\frac{a}{s+a}\right)^k$$

and the density $F_n'(x)$ of

$$S_n = X_1 + X_2 + \ldots + X_n$$

has the L.T.

$$\left(\frac{a}{s+a}\right)^{nk}.$$

Thus

$$F_n'(x) = \frac{a^{nk}\, x^{nk-1}\, e^{-ax}}{(nk-1)!}$$

and hence

$$F_n(x) = \int_0^x F_n'(y)\, dy$$

$$= 1 - e^{-ax} \sum_{r=0}^{nk-1} \frac{(ax)^r}{r!}, \quad n \geqslant 1.$$

Thus

$$p_n(t) = F_n(t) - F_{n+1}(t)$$

$$= e^{-at} \sum_{r=nk}^{(n+1)k-1} \frac{(ax)^r}{r!},$$

and using (1.5), we get

$$M^*(s) = \frac{a^k}{s\,[(s+a)^k - a^k]}. \tag{1.7}$$

Markovian case: In particular, when $k = 1$, X_n has negative exponential distribution and the renewal process then reduces to a Poisson process. As is to be expected, we have then,

$$p_n(t) = e^{-at} \frac{(at)^n}{n!}$$

and $$M^*(s) = a/s^2 \quad \text{so that} \quad M(t) = at.$$

Example 1(b). *Mixture of exponential distributions.* Let X_n have density

$$f(t) = pae^{-at} + (1-p)\, be^{-bt},\ 0 \leqslant p \leqslant 1, \quad a > b > 0. \tag{1.8}$$

Such a model may be used to describe a system which has two kinds of components—a proportion 'p' of components having a high failure rate a and the remaining proportion $(1-p)$ of components having a lower failure rate b. We have

$$f^*(s) = \frac{pa}{s+a} + \frac{(1-p)\,b}{s+b} \tag{1.9}$$

so that

$$M^*(s) = \frac{ab + s\,[pa + (1-p)\,b]}{s^2\,[s + (1-p)\,a + pb]}.$$

Writing $A = pa + (1-p)\,b$ and $B = (1-p)\,a + pb$, we get

$$\begin{aligned} M^*(s) &= \frac{As + ab}{s^2\,(s+B)} \\ &= \frac{A}{s\,(s+B)} + \frac{ab}{s^2\,(s+B)} \\ &= \frac{A}{B}\left\{\frac{1}{s} - \frac{1}{s+B}\right\} + \frac{ab}{B}\left[\frac{1}{s^2} - \left\{\frac{1}{s} - \frac{1}{s+B}\right\}\Big/B\right]. \end{aligned}$$

Inverting the L.T., we get

$$\begin{aligned} M(t) &= \frac{A}{B}\,(1 - e^{-Bt}) + \frac{ab}{B}\left\{t - \frac{1 - e^{-Bt}}{B}\right\} \\ &= \frac{abt}{B} + C(1 - e^{-Bt}), \end{aligned} \tag{1.10}$$

where

$$\begin{aligned} C &= \left(\frac{A}{B} - \frac{ab}{B^2}\right) = \frac{AB - ab}{B^2} \\ &= \frac{p\,(1-p)\,(a-b)^2}{B^2} \geqslant 0. \end{aligned}$$

Markovian case: When $p = 1$ (or $p = 0$) the distribution of X_n reduces to negative exponential and then $C = 0$, i.e. the second term of (1.10) vanishes, so that we get $M(t) = at$ (or bt).

6.2 Renewal Equation

An integral equation can be obtained for the renewal function

$$M(t) = E\,\{N(t)\},$$

which gives the expected number of renewals in $[0, t]$.

Theorem 6.2. The renewal function M satisfies the equation

$$M(t) = F(t) + \int_0^t M\,(t-x)\,dF\,(x). \tag{2.1}$$

Proof: By conditioning on the duration of the first renewal X_1, we get

$$M(t) = E\{N(t)\} = \int_0^\infty E\{N(t) \mid X_1 = x\}\, dF(x).$$

Consider $x > t$; given that $X_1 = x > t$, no renewal occurs in $[0, t]$, so that

$$E\{N(t) \mid X_1 = x\} = 0.$$

Consider $0 \leqslant x \leqslant t$; given that the first renewal occurs at x ($\leqslant t$), then the process starts again at epoch x, and the expected number of renewals in the remaining interval of length $(t - x)$ is $E\{N(t - x)\}$, so that

$$\begin{aligned} E\{N(t) \mid X_1 = x\} &= 1 + E\{N(t - x)\} \\ &= 1 + M(t - x). \end{aligned}$$

Thus considering the above two equations, we get

$$\begin{aligned} M(t) &= \int_0^t \{1 + M(t - x)\}\, dF(x) \\ &= F(t) + \int_0^t M(t - x)\, dF(x). \end{aligned}$$

We have thus established (2.1). ▲

The equation (2.1) is called the *integral equation of renewal theory* (or simply *renewal equation*) and the argument used to derive it is known as '*renewal argument*'. The renewal equation is also expressed as

$$M = F + M*f.$$

The equation (2.1) can also be established as given below.

We have
$$\begin{aligned} M(t) &= \sum_{n=1}^\infty F_n(t) = F_1(t) + \sum_{n=1}^\infty F_{n+1}(t) \\ &= F(t) + \sum_{n=1}^\infty \left\{\int_0^t F_n(t - x)\, dF(x)\right\}, \end{aligned}$$

F_{n+1} being the convolution of F_n and $F_1 = F$. Thus

$$M(t) = F(t) + \int_0^t \{\sum_{n=1}^\infty F_n(t - x)\}\, dF(x);$$

assuming the validity of the change of order of integration and summation, we get

$$M(t) = F(t) + \int_0^t M(t - x)\, dF(x).$$

It follows that $M(t) = \sum_{n=1}^{\infty} F_n(t)$ satisfies the integral equation (2.1).

The renewal equation (2.1) can be generalised as follows

$$v(t) = g(t) + \int_0^t v(t-x)\, dF(x), \quad t \geqslant 0, \tag{2.2}$$

where g and F are known and v is unknown. The equation (2.2) is called a *renewal type equation.*

A unique solution of $v(t)$ exists in terms of g and F as can be seen from the following:

Theorem 6.3. If

$$v(t) = g(t) + \int_0^t v(t-x)\, dF(x), \; t \geqslant 0 \tag{2.2}$$

then
$$v(t) = g(t) + \int_0^t g(t-x)\, dM(x), \tag{2.3}$$

where
$$M(t) = \sum_{n=1}^{\infty} F_n(t).$$

Proof: Taking L.T. of (2.2), we get

$$v^*(s) = g^*(s) + v^*(s) f^*(s)$$

or,
$$v^*(s) = \frac{g^*(s)}{1 - f^*(s)}$$

$$= g^*(s) \left[1 + \frac{f^*(s)}{1 - f^*(s)} \right]$$

$$= g^*(s) [1 + sM^*(s)].$$

Inverting the L.T., we get

$$v(t) = g(t) + \int_0^t g(t-x)\, dM(x)$$

and the solution $v(t)$ is unique, since a function is uniquely determined by its L.T. Thus the theorem is proved. ▲

For examples of phenomena satisfying a renewal type equation, see Feller II (Ch. XIV. 2).

6.3 Renewal Theorems

For the Poisson process which is a renewal process with exponential interarrival times X_n,

$$M(t) = at$$

or,
$$M(t)/t = a = 1/E(X_n).$$

In the general case, the result

$$M(t)/t \to 1/\mu, \quad \mu = E(X_n) \leqslant \infty$$

holds as $t \to \infty$. The result (due to Feller) is known as elementary renewal theorem. Before considering this, we consider a simple result.

Theorem 6.4. With probability 1,

$$\frac{N(t)}{t} \to \frac{1}{\mu} \text{ as } t \to \infty, \tag{3.1}$$

where $\mu = E(X_n) \leqslant \infty$.

Proof: Consider an interval $[0, t]$, we have

$$S_{N(t)} \leqslant t < S_{N(t)+1} \tag{3.2}$$

Now the strong law of large numbers holds for the sequence $\{S_n\}$ so that, as $n \to \infty$

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n} \to E(X_n) = \mu$$

with probability 1. Again

$$\text{as } t \to \infty, N(t) \to \infty$$

with probability 1. Thus, with probability 1,

$$\frac{S_{N(t)}}{N(t)} \to \mu \text{ as } t \to \infty. \tag{3.3}$$

Similarly, with probability 1,

$$\frac{S_{N(t)+1}}{N(t)} = \frac{S_{N(t)+1}}{N(t)+1} \cdot \frac{N(t)+1}{N(t)} \to \mu \text{ as } t \to \infty. \tag{3.4}$$

Thus from the three relations, we get that, with probability 1,

$$N(t)/t \to 1/\mu \text{ as } t \to \infty. \blacktriangle$$

This theorem shows that for large t, the average number of renewals per unit time converges to $1/\mu$.

We now go to the renewal theorems. The elementary renewal theorem is due to Feller (1941); here we shall follow the arguments put forward by Smith (1958).

Wald's Equation: Before going to renewal theorems, we discuss a special type of non-negative r.v. associated with a sequence of r.v. $\{X_i\}$ (or a stochastic process $\{X(t)\}$). Such a variable, first considered by Wald (1947) while formulating sequential analysis, is known as a *r.v. independent of the future* or a *Markov time* or a *stopping time*. For an elaborate discussion on Markov time and its applications, see A.N. Sirjaev (1973). We consider a somewhat special case here.

Definition. An integer-valued random varible N is said to be a *stopping time* for the sequence $\{X_i\}$ if the event $\{N = n\}$ is independent of

$$X_{n+1}, X_{n+2}, X_{n+3}, \ldots \text{ for all } n = 1, 2, \ldots.$$

It can be shown that N is a proper random variable.

As an example of such a variable, consider a coin tossing experiment and the result X_i of the ith toss with $X_i = 1$ or 0 depending on the result being head or tail respectively.

The sum

$$S_n = X_1 + \ldots + X_n$$

denotes the number of heads in the first n toss. Suppose that m is a given integer, then $N = \min \{n \text{ such that } S_n = m\}$ is a stopping time.

Consider the number $N(t)$ of renewals by time t, w.r.t. a sequence of interarrival times $\{X_i\}$. Now $N(t) = n$ whenever $S_n \leqslant t$ and $S_{n+1} > t$, i.e. the event $N(t) = n$ depends not only on $X_1, \ldots, X_n$ but also on X_{n+1}. Consider the variable $N(t) + 1$. Now $N(t) + 1 = n$ implies $S_{n-1} \leqslant t$ and $S_n > t$, so that the event $\{N(t) + 1 = n\}$ is independent of $X_{n+1}, X_{n+2}, \ldots$. Thus $N(t) + 1$ is a stopping time for the sequence $\{X_i\}$ while $N(t)$ is not.

From a corollary to Theorem 1.3 we find that: if

$$S_N = \sum_{i=1}^{N} X_i,$$

where X_i are i.i.d. random variables and N is a r.v. (independent of X_i's) having finite expectation, then

$$E\{\sum_{i=1}^{N} X_i\} = E(X_i)\, E(N).$$

The same result holds also when N is a stopping time for the sequence $\{X_i\}$. The proposition which is used in proving certain results in renewal theory is given below.

Wald's equation : If $\{X_i\}$ is a sequence of i.i.d. random variables and N is a stopping time for $\{X_i\}$ and $E(N) < \infty$, then

$$E\{\sum_{i=1}^{N} X_i\} = E\{X_i\}\, E\{N\}.$$

Proof : Let
$$Z_i = \begin{cases} 1 & \text{if} \quad N \geqslant i \\ 0 & \text{if} \quad N < i, \end{cases}$$

so that
$$\sum_{i=1}^{N} X_i = \sum_{i=1}^{\infty} X_i Z_i;$$

and thus
$$E\{\sum_{i=1}^{N} X_i\} = E\{\sum_{i=1}^{\infty} X_i Z_i\} = \sum_{i=1}^{\infty} E\{X_i Z_i\},$$

assuming the validity of the change of order of expectation and summation. Now Z_i is determined by $\{N < i\}$, i.e. by $X_1, \ldots, X_{i-1}$ and is thus independent of X_i. Thus

$$E\{\sum_{i=1}^{N} X_i\} = \sum_{i=1}^{\infty} E\{X_i\}\, E\{Z_i\}$$

$$\begin{aligned}&= E\{X_i\} \sum_{i=1}^{\infty} E\{Z_i\} \\ &= E\{X_i\} \sum_{i=1}^{\infty} \Pr\{N \geqslant i\} \\ &= E\{X_i\} E\{N\}.\end{aligned}$$

Hence proved. ▲

Corollary : Since $N(t)+1$ is a stopping time for the sequence $\{X_i\}$, we have,

$$\begin{aligned}E\{S_{N(t)+1}\} &= E\{\sum_{i=1}^{N(t)+1} X_i\} \\ &= E\{X_i\}\, E\{N(t)+1\} \\ &= E\{X_i\}.\{M(t)+1\}.\end{aligned}$$

Theorem 6.5. *The Elementary Renewal Theorem.* We have

$$\frac{M(t)}{t} \to \frac{1}{\mu} \quad \text{as} \quad t \to \infty, \tag{3.5}$$

where $\mu = E(X_n) < \infty$, the limit being interpreted as 0 when $\mu = \infty$.

Proof: Let

$$N = N(t)+1 \quad \text{and} \quad S_N = t + Y(t),$$

where $Y(t)$ is the residual lifetime of the unit in use at time t. Let $\{X_i^{(j)}, j = 1; 2, \ldots\}$, be a sequence of independent realisations of the renewal process $\{X_i\}$, $S_n^{(j)}$ be the corresponding partial sums, and $N^{(j)}(t)$, the corresponding number of renewals in $[0, t]$.

Let
$$\tau = S_N,\ M_k = N^{(1)}(t) + \ldots + N^{(k)}(t),$$
$$T_k = \tau^{(1)} + \ldots + \tau^{(k)}.$$

Now T_k is the sum of $(k+M_k)$ i.i.d. random variables X_i. Thus, by the strong law of large numbers, if $\mu = E(X_i) < \infty$, then as $k \to \infty$

$$\frac{T_k}{k+M_k} = \frac{\sum_{i=1}^{k+M_k} X_i}{k+M_k} \to \mu \tag{3.6}$$

with probability 1.

By the same law, we get, as $k \to \infty$,

$$\frac{M_k}{k} = \frac{\sum_{j=1}^{k} N^{(j)}(t)}{k} \to E\{N(t)\} = M(t) \tag{3.7}$$

and

$$\begin{aligned}\frac{T_k}{k} = \frac{\sum_{j=1}^{k} \tau^{(j)}}{k} &\to E(\tau) = E\{t + Y(t)\} \\ &= t + E\{Y(t)\}\end{aligned} \tag{3.8}$$

with probability 1. Combining (3.6), (3.7) and (3.8), we get, as $k \to \infty$

$$\frac{T_k}{k+M_k} = \frac{T_k}{k}\left[1+\frac{M_k}{k}\right]^{-1} \to \frac{t+E(Y(t))}{1+M(t)} \to \mu$$

or,
$$t + E(Y(t)) = \mu\,\{1 + M(t)\}$$

i.e.
$$1 + \frac{E(Y(t))}{t} = \mu\left\{\frac{1}{t} + \frac{M(t)}{t}\right\};$$

and now $Y(t)$ is positive, $E(Y(t))$ finite and hence as $t \to \infty$

$$\lim_{t\to\infty}\inf \frac{M(t)}{t} = \frac{1}{\mu}. \tag{3.9}$$

To prove the theorem we have to show that

$$\lim_{t\to\infty}\sup \frac{M(t)}{t} \leqslant \frac{1}{\mu}.$$

For doing this we define a new renewal process as follows:

Let A be a constant > 0 and for $n = 1, 2, \ldots$, and let

$$X_n^* = \begin{cases} X_n & \text{if } X_n \leqslant A \\ A & \text{if } X_n > A \end{cases}$$

Let
$$S_n^* = \sum_{i=1}^{n} X_i^* \text{ and } N^*(t) = \sup\{n : S_n^* \leqslant t\},$$

$$M^*(t) = E\{N^*(t)\}.$$

Now
$$S_n^* \leqslant S_n, \text{ and hence } N^*(t) \geqslant N(t)$$

and
$$M^*(t) \geqslant M(t).$$

Again
$$E(X_n^*) = \mu_A \leqslant \mu \text{ and } \mu_A \to \mu, \text{ as } A \to \infty.$$

We have
$$S^*_{N(t)+1} \leqslant t + A$$

$$E\,(S^*_{N(t)+1}) = \mu_A\,(M^*(t) + 1)$$

and hence
$$\lim_{t\to\infty}\sup \frac{M^*(t)}{t} \leqslant \frac{1}{\mu_A}$$

and as $A \to \infty$,
$$\lim_{t\to\infty}\sup \frac{M(t)}{t} \leqslant \frac{1}{\mu}. \tag{3.10}$$

From (3.9) and (3.10) the result follows. ▲

Example 3(a). *Age and block replacement policies:*
(Barlow and Proschan, 1964)

The usual replacement policy implies replacement of a component, as and when it fails, by a similar new one. There are other policies besides this; the two most important replacement policies that are in use, in general, are the age and block replacement policies. Under an *age replacement policy*, a component is replaced upon failure *or* when it attains a specified age T, whichever occurs earlier. Under a *block replacement policy*, a component is replaced upon failure *and* also regularly at times $T, 2T, \ldots$. We shall call T, the *replacement interval* under these two policies. We assume, as usual, that components or units fail permanently, independently and that the detection of a failure and replacement of the failed item are instantaneous. Suppose that the successive lifetimes of the units are random variables with a common distribution function $F(x)$ having mean μ. Considering failures as renewals, the number of failures in $[0, t]$ under the three replacement policies, the usual one, the age replacement and the block replacement can be denoted by three renewal processes as indicated below:

$N(t)$ = the number of failures in $[0, t]$ for an ordinary renewal process.

$N_A(t, T)$ = the number of failures in $[0, t]$ under age replacement policy, with replacement interval T.

$N_B(t, T)$ = the number of failures in $[0, t]$ under block replacement policy, with replacement interval T.

Let the corresponding renewal functions be

$$\begin{aligned} M(t) &= E\{N(t)\} \\ M_A(t, T) &= E\{N_A(t, T)\} \\ M_B(t, T) &= E\{N_B(t, T)\}. \end{aligned} \tag{3.11}$$

It will be shown here (see Barlow and Proschan) that

$$\lim_{t\to\infty} \frac{N_A(t, T)}{t} = \lim_{t\to\infty} \frac{M_A(t, T)}{t} = \frac{F(T)}{\int_0^T \{1 - F(x)\}\, dx} \tag{3.12}$$

$$\lim_{t\to\infty} \frac{N_B(t, T)}{t} = \lim_{t\to\infty} \frac{M_B(t, T)}{t} = \frac{M(T)}{T}. \tag{3.13}$$

To prove (3.12) we first note that $\{N_A(t, T), t \geqslant 0\}$ is a renewal process with distribution

$$F_A(t) = \Pr(Y_i \leqslant t)$$

such that, in

$$nT \leqslant t < (n+1)\, T, n = 0, 1, 2, \ldots,$$

$$1 - F_A(t) = \Pr(Y_i > t) = [1 - F(T)]^n\, [1 - F(t - nT)]. \tag{3.14}$$

Hence

$$\begin{aligned} \mu_A = E(Y_i) &= \int_0^\infty \Pr(Y_i \geqslant t)\, dt \\ &= \sum_{n=0}^{\infty} \int_{nT}^{(n+1)T} \{[1 - F(T)]^n\, [1 - F(t - nT)]\}\, dt \end{aligned}$$

$$= \sum_{n=0}^{\infty} [1 - F(T)]^n \left[\int_0^T \{1 - F(x)\}\, dx \right]$$

$$= \int_0^T \{1 - F(x)\}\, dx / F(T). \tag{3.15}$$

With this value of μ_A, using Theorems 6.4 and 6.5, we at once get (3.12).

To prove (3.13) we note that after successive block replacements at $T, 2T, \ldots$, the process starts anew, and if $N_{B_r}(T)$ is the number of failures in the interval $[(r-1)T, rT]$, $r = 1, 2, 3, \ldots$, following block replacements, then $N_{B_r}(T) = N(T)$.

Suppose that for some specified integral value of n, $nT \leq t < (n+1)T$, i.e. for some n and $0 \leq \tau < T$,

$$t = nT + \tau,$$

then

$$N_B(t, T) = \sum_{r=1}^{n} N_{B_r}(T) + N(\tau)$$

$$= \sum_{r=1}^{n} N(T) + N(\tau).$$

We then get

$$\lim_{t\to\infty} \frac{N_B(t, T)}{t} = \lim_{n\to\infty} \frac{\sum_{r=1}^{n} N(t)}{nT}$$

$$= E\{N(T)\}/T, \text{ (by the law of Large Numbers)}$$

$$= M(T)/T.$$

Thus we get (3.13). ▲

Particular Cases

(i) *Markovian Case*: When $F(x) = 1 - e^{-x/a}$ with mean $\mu = a$, then

$$\mu_A = \int_0^T [1 - F(x)]\, dx / F(T) = a = \mu$$

and $\quad M(T)/T = aT/T = a = \mu,$

as is to be expected (Explain why?).

(ii) *Limiting Case*: Suppose that μ is finite, then for large T, we have asymptotically,

$$1/\mu_A \to 1/\mu$$

and by Theorem 6.5, $M(T)/T \to 1/\mu$.

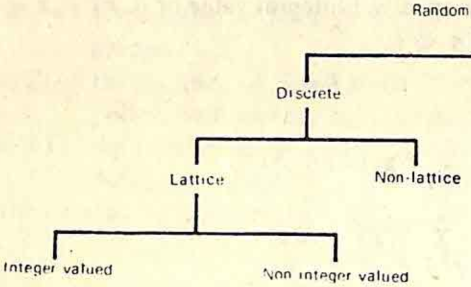
It should be intuitively clear that for $T \to \infty$, the renewal processes under age and block replacement policies are both equivalent to the ordinary renewal process $N(t)$. For some interesting properties and renewal theoretical implications of these replacement policies, refer to Barlow and Proschan (1964, 1975).

We shall state without proof two other renewal theorems hereafter.

Definition. A discrete random variable X is said to be a *lattice variable* (or to have a *lattice distribution*) if it takes values $c + nd$ ($n = 0, \pm 1, \pm 2, \ldots$), where c and d (>0) are constants. The largest d is said to be the *period* of the distribution. (Feller uses the terminology 'arithmetic' for lattice and 'span' for period but the terminology used here is more common.)

Lattice variables form a special type of the class of discrete variables. When $c = 0$, $d = 1$, a lattice variable becomes an integer-valued variable.

The classification of random variables is given below:*



By saying that a distribution is *not lattice*, we shall mean that the corresponding random variable is other than lattice, that is, it is either discrete non-lattice or continuous.

Theorem 6.6. (Blackwell's theorem): For X_i not lattice and for fixed $h \geqslant 0$

$$M(t) - M(t-h) \to h/\mu \quad \text{as } t \to \infty$$

and for a lattice X_i with period d,

$$\lim_{n \to \infty} \Pr\{\text{renewal at } nd\} \to d/\mu.$$

A more general form of the theorem is given below.

Theorem 6.7. (Smith's theorem or the Key Renewal theorem):

Let $h(t)$ be a non-negative, non-increasing function of $t \geqslant 0$ such that

$$\int_0^\infty h(t)\, dt < \infty$$

and X_i be not lattice. Then

$$\text{as } t \to \infty \qquad \int_0^t h(t-x)\, dM(x) \to \frac{1}{\mu} \int_0^t h(t)\, dt,$$

the limit being interpreted as 0 when $\mu = \infty$.

---

* A r.v. may also be of the mixed type, being partly discrete and partly continuous.

For proof of Blackwell's and Smith's theorems, see Prabhu (1965), Feller (1966).

Smith's theorem is called the key renewal theorem as a number of important results can be derived from it. For example, if

$$\begin{aligned} h(t) &= 1/h, \quad 0 < t \leq h \\ &= 0, \quad \text{otherwise;} \end{aligned}$$

we at once get Blackwell's theorem

$$M(t) - M(t-h) \to \frac{1}{\mu}\int_0^h dt = h/\mu.$$

6.4 Delayed and Equilibrium Renewal Processes

The renewal process $\{X_i\}$ which we have considered so far is such that all the X_i's, $i = 1, 2, \ldots$ are i.i.d r.v.'s The process is what is known as *ordinary renewal process*. There are two generalisations of the ordinary renewal process.

First, suppose that the first interarrival time X_1 (i.e. time from the origin upto the first renewal) has a distribution G which is different from the common distribution F of the remaining interarrival times $X_2, X_3, \ldots$, i.e. the initial distribution G is different from subsequent common distribution F. We then get what is known as a *modified* or *delayed renewal process*. Such a situation arises when the component used at $t = 0$ is not new. When $G = F$, the modified renewal process reduces to the ordinary renewal process.

Definition. Let $\{X_i, i = 1, 2, \ldots\}$ be a sequence of independent non-negative random variables such that the first interarrival time X_1 has the distribution G and the subsequent interarrival times X_n, $n = 2, 3, \ldots$ have the identical distribution F.

Let

$$S_0 = 0,\ S_n = \sum_{i=1}^{n} X_i \text{ and } N_D(t) = \sup\{n: S_n \leq t\}.$$

The stochastic process $\{N_D(t), t \geq 0\}$ is called a *modified*, *delayed* or a *general renewal process*.

The sequence of random variables $\{X_n, n = 1, 2, \ldots\}$, where the initial distribution G is different from subsequent common distribution F, is said to constitute a modified renewal process.

The following results can be easily obtained:

We have

$$\begin{aligned} \Pr\{N_D(t) = n\} &= \Pr\{S_n \leq t\} - \Pr\{S_{n+1} \leq t\} \\ &= G * F_{n-1}(t) - G * F_n(t) \end{aligned}$$

so that

$$M_D(t) = E\{N_D(t)\} = \sum_{n=1}^{\infty} G * F_{n-1}(t) \tag{4.1}$$

and

$$M_D^*(s) = \frac{g^*(s)}{s\{1 - f^*(s)\}}, \tag{4.2}$$

g^* and f^* being the L.T. of the p.d.f. of G and F respectively.

Again,

$$\begin{aligned} M_D(t) &= \int_0^\infty E\{N_D(t) \mid X_1 = x\}\, dG(x) \\ &= \int_0^t \{1 + M(t-x)\}\, dG(x) \\ &= G(t) + \int_0^t M(t-x)\, dG(x); \end{aligned} \tag{4.3}$$

$M(t)$ being the renewal function of the ordinary renewal process with distribution F.

Example 4(a). Consider a modified renewal process with the first interarrival time X_1 having mixed exponential distribution (considered in Example 1(b) and the other interarrival times $X_2, X_3, \ldots$ having simple exponential distribution with parameter b. Such a model would be suitable when there is no defective item among replacement parts (i.e. whereas X_1 has distribution as described therein, the distribution of $X_2, X_3, \ldots$ corresponds to $p = 0$ in the distribution of X_1. Such a situation has been visualised by Karmarkar, 1978). We have

$$g(t) = pae^{-at} + (1-p)\, be^{-bt},\ a > b > 0.$$

From (4.2) we get

$$\begin{aligned} M_D^*(s) &= \frac{g^*(s)}{s\{1 - f^*(s)\}} \\ &= \frac{ab + \{b + p\,(a-b)\}\, s}{s^2\,(s+a)} \\ &= \frac{ab}{s^2\,(s+a)} + \frac{b + p\,(a-b)}{s\,(s+a)}. \end{aligned}$$

Inverting, we get

$$\begin{aligned} M_D(t) &= \frac{ab}{a}\left\{t - \frac{1 - e^{-at}}{a}\right\} + \frac{b + p\,(a-b)}{a}(1 - e^{-at}) \\ &= bt + \frac{p(a-b)}{a}(1 - e^{-at}). \end{aligned}$$

$M_D(t)$ can also be obtained directly from (4.3), by noting that the renewal function $M(t)$ of the ordinary renewal process with distribution F is given by

$$M(t) = bt$$

and
$$G(t) = \int_0^t g(x)\, dx.$$

We state, without proof, the corresponding renewal theorem for modified renewal process.

Theorem 6.8. Let $\mu = E\{X_n\}$, $n = 2, 3, \ldots$. Then

(1) with probability 1,

$$M_D(t)/t \to 1/\mu \text{ as } t \to \infty. \tag{4.4}$$

(2) If F is not lattice, then

$$M_D(t) - M_D(t-a) \to a/\mu \text{ as } t \to \infty \tag{4.5}$$

and if F is lattice with period d, then

$$\Pr\{\text{renewal at } nd\} \to d/\mu \text{ as } n \to \infty.$$

The above theorem shows that the renewal rate tends to a constant and is independent of the initial distribution G. We have, as Feller observes, an analogue to the Ergodic theorem for Markov chains.

Now $M_D(t) \to t/\mu$ raises a question: whether there exists an initial distribution G such that $M_D(t) = t/\mu$. The answer is in the affirmative. This is true in case of a kind of renewal process discussed below.

Equilibrium (or stationary) Renewal Process

Now a modified renewal process with initial distribution $G = F_e$ exists such that

$$M_e(t) = t/\mu,$$

or
$$M_e^*(s) = 1/\mu s^2$$

iff
$$f_e^*(s) = \frac{1 - f^*(s)}{\mu s},$$

or
$$f_e(t) = \frac{1 - F(x)}{\mu},$$

i.e.
$$F_e(t) = \int_0^t \frac{1 - F(x)}{\mu}\, dx. \tag{4.6}$$

The distribution having d.f. $F_e(t)$ is called equilibrium distribution of F and the corresponding renewal process is called equilibrium (or stationary) renewal process. We denote an equilibrium process by $\{N_e(t),\ t \geq 0\}$ and the corresponding renewal function by $M_e(t)$. Poisson process is clearly an equilibrium renewal process.

Equilibrium renewal processes, which are extremely important in applications, can arise as follows: Suppose that an ordinary renewal process, has been continuing from time $t \to -\infty$, remote from $t = 0$. If the observation of the process is started at $t = 0$, then the interarrival time

X_1 to the first renewal will have distribution function $G = F_e$ as given by (4.6).

Example 4(b). We have, for all $t > 0$, and $n = 0, 1, 2, \ldots$

$$\Pr\{N_e(t) \geqslant n\} \geqslant \Pr\{N(t) \geqslant n\},$$

iff

$$\frac{\int_t^\infty \{1 - F(x)\}\,dx}{1 - F(t)} \leqslant \mu.$$

Proof: Assume that the condition holds. Since

$$\frac{1}{\mu}\int_t^\infty \{1 - F(x)\}\,dx = \frac{1}{\mu}\left[\int_0^\infty \{1 - F(x)\}\,dx - \int_0^t \{1 - F(x)\}\,dx\right]$$

$$= 1 - F_e(t),$$

we have by hypothesis, for $t \geqslant 0$

$$1 - F_e(t) \leqslant 1 - F(t)$$

or,
$$F_e(t) \geqslant F(t).$$

If S_n and S_n' denote the nth renewal epochs of $N(t)$ and $N_e(t)$ respectively, then, for $n \geqslant 1$,

$$\begin{aligned}\Pr\{N_e(t) \geqslant n\} &= \Pr\{S_n' \leqslant t\}\\ &= \int_0^t F^{n-1}(t - x)\,dF_e(x)\\ &\geqslant \int_0^t F^{n-1}(t - x)\,dF(x)\\ &= \Pr\{S_n \leqslant t\} = \Pr\{N(t) \geqslant n\}.\end{aligned}$$

Further, $\Pr\{N_e(t) \geqslant 0\} = \Pr\{N(t) \geqslant 0\} = 1$, for all $t \geqslant 0$. Thus the result holds.

To prove the converse, assume that $\Pr\{N_e(t) \geqslant n\} \geqslant \Pr\{N(t) \geqslant n\}$ holds for $n = 0, 1, 2, \ldots$. Since equality holds for $n = 0$, we must have

$$\Pr\{N_e(t) = 0\} \leqslant \Pr\{N(t) = 0\}$$

or,
$$1 - F_e(t) \leqslant 1 - F(t)$$

and therefore the converse is also true. ▲

Corollary. $M(t) \leqslant t/\mu$ iff $F_e(t) \geqslant F(t)$.

Note: We describe two concepts used to classify distributions in reliability theory.

A unit (or equivalently its lifetime distribution F) is said to be New Better (Worse) than Used NBU (NWU) if

$$\{1 - F(x + y)\} \leqslant (\geqslant) [\{1 - F(x)\} \{1 - F(y)\}].$$

In other words, a unit is NBU (NBW) if the survival probability of a unit at age x is less (greater) than the corresponding survival probability of a new unit.

A unit is said to be New Better (Worse) than Used in Expectation NBUE (NWUE) iff

$$\frac{\int_t^\infty \{1 - F(x)\}\, dx}{1 - F(t)} \leqslant (\geqslant) \mu.$$

In other words, a unit is NBUE (NWUE) if the conditional expectation of the residual life of a unit at age t is less (greater) than the expectation of the lifetime of a new unit.

The result given in the above example may thus be restated as:

$$\Pr\{N_e(t) \geqslant n\} \geqslant \Pr\{N(t) \geqslant n\}, \quad t \geqslant 0, n = 0, 1, 2, \ldots$$

as well as $M(t) \to t/\mu$, iff the distribution F inducing the renewal process is NBUE.

Probability Generating Function (p.g.f.) of Renewal Processes:
We obtain here the p.g.f. of the renewal processes

$$\{N(t), t \geqslant 0\}, \{N_D(t), t \geqslant 0\}, \{N_e(t), t \geqslant 0\}.$$

Let

$$\begin{aligned} G(t, z) &= \sum_{r=0}^{\infty} z^r \Pr\{N(t) = r\} \\ &= \sum_{r=0}^{\infty} z^r \{F_r(t) - F_{r+1}(t)\} \\ &= 1 + \sum_{r=1}^{\infty} z^{r-1} (z - 1) F_r(t). \end{aligned} \tag{4.7}$$

The above relation holds for all the three types of renewal processes. We denote the p.g.f. of ordinary, delayed (modified) and equilibrium renewal processes by G, G_m and G_e respectively. Now the Laplace transform of the p.d.f. of F_r (for $r \geqslant 1$) is

$$f_r^*(s) = [f^*(s)]^r \quad \text{(for an ordinary renewal process)} \tag{4.8a}$$

$$= g^*(s)[f^*(s)]^{r-1} \quad \text{(for a delayed renewal process)} \tag{4.8b}$$

$$= \frac{1 - f^*(s)}{\mu s} [f^*(s)]^{r-1} \quad \text{(for an equilibrium renewal process).} \tag{4.8c}$$

Denote the L.T. of the p.g.f. of the ordinary, modified and equilibrium renewal processes by

$$G^*(s, z), G_m^*(s, z), G_e^*(s, z)$$

respectively. From (4.7), we have

$$G^*(s, z) = \frac{1}{s} + \frac{1}{s}\sum_{r=1}^{\infty} z^{r-1}(z-1)\,[f^*(s)]^r$$

$$= \frac{1}{s} + \frac{1}{s}\cdot\frac{(z-1)\,f^*(s)}{1-z\,f^*(s)}$$

$$= \frac{1-f^*(s)}{s\,\{1-zf^*(s)\}}, \tag{4.9}$$

$$G_m^*(s, z) = \frac{1}{s} + \frac{1}{s}\sum_{r=1}^{\infty} z^{r-1}(z-1)\,g^*(s)\,[f^*(s)]^{r-1}$$

$$= \frac{1}{s} + \frac{(z-1)\,g^*(s)}{s\,\{1-z\,f^*(s)\}}$$

$$= \frac{1-z\,f^*(s)+zg^*(s)-g^*(s)}{s\,\{1-z\,f^*(s)\}}, \tag{4.10}$$

$$G_e^*(s, z) = \frac{1}{s} + \frac{1}{s}\sum_{r=1}^{\infty} z^{r-1}(z-1)\,\frac{1-f^*(s)}{\mu s}\,[f^*(s)]^{r-1}$$

$$= \frac{1}{s} + \frac{(z-1)\{1-f^*(s)\}}{\mu s^2\,\{1-z\,f^*(s)\}} \tag{4.11}$$

$$= \frac{1}{s} + \frac{(z-1)}{\mu s}\,G^*(s, z). \tag{4.11a}$$

Inverting (4.11a), we get

$$G_e(t, z) = 1 + \frac{z-1}{\mu}\int_0^t G\,(u, z)\,du;$$

comparing the coefficients of z^r from both sides, we get

$$\Pr\{N_e(t) = r\} = \frac{1}{\mu}\int_0^t [\Pr\{N(u) = r-1\} - \Pr\{N(u) = r\}]\,du \tag{4.12}$$

$$\Pr\{N_e(t) = 0\} = 1 - \frac{1}{\mu}\int_0^t \Pr\{N(u) = 0\}\,du. \tag{4.12a}$$

Example 4(c). *Second moment and variance of $N(t)$.*

The moments of $N(t)$ can be found from the L.T. of the p.g.f. However, the following simple procedure may be employed to find the second moment of $N(t)$ and thence its variance. We have

$$L(t) = E\,\{N(t)\}^2 = \sum_{r=0}^{\infty} r^2\,\{F_r(t) - F_{r+1}(t)\}.$$

Taking L.T., we get

$$L^*(s) = \sum_{r=0}^{\infty} \frac{r^2}{s}\,[\{f^*(s)\}^r - \{f^*(s)\}^{r+1}].$$

On simplification, we get

$$L^*(s) = \frac{f^*(s)\{1+f^*(s)\}}{s\{1-f^*(s)\}^2}$$

$$= \left[\frac{1+f^*(s)}{1-f^*(s)}\right] M^*(s)$$

which can also be written as:

$$L^*(s) = M^*(s) + \frac{(2f^*(s))}{s\{1-f^*(s)\}}(sM^*(s))$$

Inverting the L.T., we get

$$E[(N(t))]^2 = L(t) = M(t) + 2\int_0^t M(t-x)\,dM(x). \tag{4.13}$$

This gives the second moment of $N(t)$ in terms of the renewal function $M(t)$. Thus var$\{(N(t)\}$ can be obtained in terms of $M(t)$.

Example 4(d). *An application to a management problem.* For certain kinds of electronic equipments of office use (such as electronic calculators etc.) and of domestic appliances (such as T.V., refrigerators etc.) the manufacturer offers to the customer a certain kind of service contract by which the manufacturer undertakes to maintain and repair the product over a certain specified period against payment of a fee by the customer. In arriving at a decision regarding the offer of the terms of contract (including the amount of fees to be charged thereof), the manufacturer should be aware of the expected cost of future repair. A simple renewal theoretic model developed by Karmarkar (1978) is considered here. It is assumed that each failure is followed by instantaneous detection and repair.

Denote

$R(t)$ = cost associated with a repair (renewal) at time t; it is assumed that $R(t) = R$,

c = the discount factor for "continuous" discounting of future costs,

$f(t)$ = the p.d.f. of time to failure (interarrival time X_n),

$k_r(t)$ = the p.d.f. of $S_r = X_1 + \dots + X_r$; F_r = d.f. of S_r,

V_r = the discounted cost of rth renewal.

For a single failure followed by renewal, the expected discounted cost $E(V_1)$ is given by

$$E(V_1) = \int_0^\infty Re^{-ct} f(t)\,dt = Rf^*(c).$$

The expected discounted cost for the rth renewal is

$$E(V_r) = \int_0^\infty Re^{-ct}\, k_r(t)\, dt$$
$$= Rk_r^*(c)$$
$$= R\,[f^*(c)]^r.$$

For a sequence of n renewals, the total expected discounted cost is

$$E(V) \equiv E(V_1 + V_2 + \ldots + V_n) = \sum_{r=1}^{n} R\,[f^*(c)]^r$$
$$= \frac{R\,[1 - \{f^*(c)\}^n]f^*}{1 - f^*(c)}$$
$$\to \frac{R\,f^*(c)}{1 - f^*(c)} \quad \text{as } n \to \infty. \tag{4.14}$$

For a sequence of infinite renewals truncated at time T the expected cost of future renewals is

$$E\,\{V(T)\} = \sum_{r=1}^{n} \int_0^T Re^{-ct}\, k_r(t)\, dt$$
$$= \int_0^T Re^{-ct}\, d\Big(\sum_{r=1}^{n} F_r(t)\Big)$$
$$= \int_0^T Re^{-ct}\, dM(t), \tag{4.15}$$

assuming the interchangeability of summation, integration and differentiation. As $T \to \infty$,

$$E\,\{V(T)\} \to RcM^*(c) = \frac{Rf^*(c)}{1 - f^*(c)} \tag{4.16}$$

which agrees with (4.14) as $n \to \infty$.

Particular cases

(1) Suppose that $f(t) = be^{-bt}$, then

$$E(V) = \frac{R\left[1 - \left(\frac{b}{c+b}\right)^n\right]\frac{b}{c+b}}{1 - \frac{b}{b+c}}$$
$$= \left(\frac{Rb}{c}\right)\left[1 - \left(\frac{b}{c+b}\right)^n\right]$$
$$\to \frac{Rb}{c} \qquad \text{as } n \to \infty$$

and for infinite renewals truncated at T, we get

$$E\{V(T)\} = \int_0^T Re^{-ct}\, d(bt)$$

$$= \left(\frac{Rb}{c}\right)(1 - e^{-cT})$$

$$\to \frac{Rb}{c} \text{ as } T \to \infty.$$

(2) Suppose that $f(t) = pae^{-at} + (1-p)\, be^{-bt}$, $0 \leqslant p \leqslant 1$, $a > b > 0$, i.e. X_n has a mixed exponential distribution as discussed in Example 1(b). Then

$$E(V) \to R\left[\frac{pa}{c+a} + \frac{(1-p)\,b}{c+b}\right] \Big/ \left[1 - \left\{\frac{pa}{c+a} + \frac{(1-p)\,b}{c+b}\right\}\right]$$

$$= \frac{R\,\{Ac + ab\}}{c\,(c+B)},$$

where $\qquad A = pa + (1-p)\,b, \quad B = (1-p)\,a + pb,$

$$E\{V(T)\} = \int_0^T Re^{-ct}\left\{\frac{ab}{B} + \left(\frac{A}{B} - \frac{ab}{B^2}\right) Be^{-Bt}\right\} dt$$

$$= \frac{Rab}{cB}\,(1 - e^{-cT}) + \frac{R(AB - ab)}{B(c+B)}\,[1 - e^{-(c+B)}].$$

(3) Suppose that

$$f(t) = \frac{\lambda^k t^{k-1}\, e^{-\lambda t}}{\Gamma(k)}, \quad t \geqslant 0$$

so that

$$f^*(s) = \left(\frac{\lambda}{s+\lambda}\right)^k;$$

then

$$E(V) \to \frac{R\lambda^k}{(c+\lambda)^k - \lambda^k}$$

Example 4(e): *Number of renewals in a random time.* So far we have considered the number $N(t)$ of renewals in a fixed interval $[0, t]$. Consider now that T is a random variable independent of X_i so that $[0, T]$ is a random interval; suppose that N is the number of renewals in $[0, T]$. Let $G(t, z)$ be the p.g.f. of $N(t)$ and $G(z)$ the p.g.f. of N. If $q(t)$ is the p.d.f. of T, then it is clear that

$$G(z) = \int_0^\infty G(t, z)\, q(t)\, dt. \tag{4.17}$$

Particular cases

(1) Suppose that T has exponential distribution, with p.d.f. $q(t) = ae^{-at}$. Then

$$G(z) = \int_0^\infty ae^{-at}\, G(t, z)\, dt$$
$$= aG^*(a, z).$$

If the renewal process is ordinary, then from (4.9) we get

$$G(z) = aG^*(a, z)$$
$$= \frac{a\{1 - f^*(a)\}}{a\{1 - zf^*(a)\}}$$
$$= \frac{1 - f^*(a)}{1 - zf^*(a)},$$

which is the p.g.f. of a geometric distribution.

We have

$$\Pr\{N = r\} = [f^*(a)]^r\,[1 - f^*(a)], \quad r = 0, 1, 2, \ldots.$$

Suppose further that the renewal process is Poisson with $f(t) = be^{-bt}$; then

$$G(z) = \frac{a}{a + b - zb}$$

and
$$\Pr\{N = r\} = \left(\frac{a}{a+b}\right)\left(\frac{b}{a+b}\right)^r, \; r = 0, 1, 2, \ldots.$$

This result was obtained in Ch. 4 (see Poisson count process).

(2) Suppose that the renewal process is Poisson but T has gamma distribution having p.d.f.

$$q(t) = a^k\, t^{k-1}\, e^{-at}/\Gamma(k), \quad t \geqslant 0;$$

T can be considered as the sum of k i.i.d. exponential distributions with p.d.f.
$$q(t) = ae^{-at}, \quad t \geqslant 0.$$

From the memoryless property of exponential distribution, we then get that the distribution of N is negative binomial (Pascal) with p.g.f.

$$[G(z)]^k = \left[\frac{a}{a + b - zb}\right]^k$$

so that

$$\Pr\{N = r\} = \binom{k + r - 1}{r}\left(\frac{a}{a+b}\right)^k\left(\frac{b}{a+b}\right)^r, r = 0, 1, 2, \ldots$$

6.5 Residual and Excess Lifetimes

We discuss here two random variables of interest in renewal theory. To a given $t > 0$, there corresponds uniquely a $N(t)$ such that

$$S_{N(t)} \leqslant t < S_{N(t)+1},$$

i.e. t falls in the interval $X_{N(t)+1}$.

(i) The *residual lifetime* of the individual alive at age t is given by the time $Y(t)$ from t to the next renewal epoch, i.e.

$$Y(t) = S_{N(t)+1} - t.$$

It is also called *forward recurrence time* at t.

(ii) The *spent lifetime* of the individual alive at age t is given by the time to t since the last renewal epoch, i.e.

$$Z(t) = t - S_{N(t)}.$$

It is also called *backward recurrence time* at t.

(iii) The length of the lifetime containing t is given by

$$Y(t) + Z(t) = S_{N(t)+1} - S_{N(t)} = X_{N(t)+1}.$$

The distributions of $Y(t)$ and $Z(t)$ are given below.

Theorem 6.9. We have

$$\Pr\{Y(t) \leqslant x\} = F(t+x) - \int_0^t [1 - F(t+x-y)]\, dM(y). \tag{5.1}$$

If in addition, F is not lattice, then

$$\lim_{t\to\infty} \Pr\{Y(t) \leqslant x\} = \frac{1}{\mu}\int_0^x [1 - F(y)]\, dy. \tag{5.2}$$

Proof: It is clear that

$$Y(t) > x \Leftrightarrow \text{no renewals in } [t, t+x].$$

Let $$P(t) = \Pr\{Y(t) > x\}.$$

By conditioning on X_1, we get

$$\begin{aligned} P(t) &= \Pr\{Y(t) > x\} \\ &= \int_0^\infty \Pr\{Y(t) > x \mid X_1 = y\}\, dF(y). \end{aligned}$$

Now three situations may arise: (i) $y > t + x$, in which case it is certain that no renewal occurs in $[t, t+x]$, so that

$$\begin{aligned} \Pr\{Y(t) > x\} &= \Pr(\text{no renewal in } [t, t+x]) \\ &= 1 \end{aligned}$$

(ii) $t < y \leqslant t + x$, in which case one renewal (the one corresponding to $X_1 = y$) occurs in $[t, t+x]$, so that

$$\Pr\{Y(t) > x\} = \Pr\{\text{no renewal in } [t, t+x]\} = 0$$

(iii) $0 < y \leqslant t$, in which case the first renewal occurs in $y, y + dy$ and then the process restarts itself and

$$\begin{aligned} \Pr\{Y(t) > x\} &= \Pr\{Y(t-y) > x\} \\ &= P(t-y). \end{aligned}$$

Considering the above three situations, we get

$$P(t) = \int_0^t P(t-y)dF(y) + \int_{t+x}^{\infty} dF(y)$$

or,
$$P(t) = 1 - F(t+x) + \int_0^t P(t-y)\, dF(y),$$

which is a renewal type equation. Applying Theorem 6.3., we have

$$P(t) = 1 - F(t+x) + \int_0^t \{1 - F(t+x-y)\}\, dM(y), \quad (5.3)$$

which is equivalent to (5.1). Thus (5.1) is proved.

Now suppose that F is not lattice. Then applying the key renewal theorem (Theorem 6.7) with $h(t) = 1 - F(t+x)$ we get, as $t \to \infty$

$$\int_0^t \{1 - F(t+x-y)\}\, dM(y) \to \frac{1}{\mu} \int_0^{\infty} \{1 - F(t+x)\}\, dt$$

$$= \frac{1}{\mu} \int_x^{\infty} [1 - F(y)]\, dy.$$

Hence from (5.3), we have

$$\lim_{t\to\infty} \Pr\{Y(t) > x\} = \lim_{t\to\infty} P(t) = \frac{1}{\mu} \int_0^{\infty} \{1 - F(y)\}\, dy$$

or,
$$\lim_{t\to\infty} \Pr\{Y(t) \leqslant x\} = 1 - \frac{1}{\mu} \int_x^{\infty} \{1 - F(y)\}\, dy$$

$$= \frac{1}{\mu} \int_0^{x} \{1 - F(y)\}\, dy.$$

Thus (5.2) is proved. ▲

Theorem 6.10. We have

$$\Pr\{Z(t) \leqslant x\} = \begin{cases} F(t) - \int_0^{t-x} \{1 - F(t-y)\}\, dM(y), & x \leqslant t; \quad (5.4) \\ 1 & x > t. \quad (5.5) \end{cases}$$

If, in addition, F is not lattice, then

$$\lim_{t\to\infty} \Pr\{Z(t) \leqslant x\} = \frac{1}{\mu} \int_0^{x} \{1 - F(y)\}\, dy.$$

Proof: The distribution of $Z(t)$ can be obtained immediately by noting that

$$Z(t) > x \Leftrightarrow \text{no renewals in } [t-x, t]$$

$$\Leftrightarrow Y(t-x) > x,$$

so that $\Pr\{Z(t) > x\} = \Pr\{Y(t-x) > x\}$,

Hence from the Theorem 6.9, we get

$$\begin{aligned}\Pr\{Z(t) \leqslant x\} &= \Pr\{Y(t-x) \leqslant x\}\\ &= F(t) - \int_0^{t-x} \{1 - F(t-y)\}\, dM(y), \quad x \leqslant t\\ &= 1 \qquad\qquad\qquad\qquad\qquad\qquad\quad x > t\end{aligned}$$

and if F is not lattice, then

$$\lim_{t\to\infty} \Pr\{Z(t) \leqslant x\} = \int_0^x \{1 - F(y)\}\, dy/\mu. \quad \blacktriangle$$

Note: It is to be noted that the result

$$\lim_{t\to\infty} \Pr\{Y(t) \leqslant x\} = \int_0^x \{1 - F(y)\}\, dy/\mu$$

holds also for the modified renewal process regardless of the initial distribution G.

Now the R.H.S. is the distribution function $F_e = G$ of the equilibrium renewal process which has a uniform renewal rate $M(t)/t = \mu$. Thus if $\mu < \infty$, then the residual lifetime $Y(t)$ has a distribution which coincides with the distribution of the process attaining uniform renewal rate.

If $\mu = \infty$, then the probability that the residual lifetime exceeds x tends to unity for all x. This should be intuitively clear.

The asymptotic behaviour of the residual lifetime distribution has been investigated at length by Balkema *et al.* (1974) who obtained the possible limit distribution types.

Example 5(a). *Markovian case*: Let X_i have exponential distribution with distribution function $F(x) = 1 - e^{-ax}$.

Then $M(t) = at$, and we have

$$\begin{aligned}\Pr\{Y(t) \leqslant x\} &= \{1 - e^{-a(t+x)}\} - \int_0^t \{1 - 1 + e^{-a(t+x-y)}\}\, d(ay)\\ &= 1 - e^{-ax} \text{ for all } t.\end{aligned}$$

Thus $Y(t)$ has the same exponential distribution. In fact this is one of the characteristics of exponential distribution.

For $x \leqslant t$,

$$\Pr\{Z(t) \leqslant x\} = 1 - e^{-at} - \int_0^{t-x} \{e^{-a(t-y)}\}\, d(ay)$$

$$= 1 - e^{-ax}$$

and for $x > t$,

$$\Pr\{Z(t) \leqslant x\} = 1.$$

For large t, $Z(t)$ will also have the same exponential distribution.

From the memoryless property of exponential distribution, it is clear that Y and Z are independent. As

$$X_{N(t)+1} = Y(t) + Z(t),$$

the renewal interval containing t is given by the convolution of the residual and spent lifetimes at t. For large t, the distribution of $X_{N(t)+1}$ will be gamma with mean $2/a$, i.e.

$$E\{X_{N(t)+1}\} = 2/a = 2E(X_i).$$

Thus for large t, the expected duration of the renewal interval containing t will be *twice* the expected duration of an ordinary renewal interval.

6.6 Renewal Reward Process

Consider an ordinary renewal process generated by the sequence

$$\{X_n,\ n = 1, 2, \ldots\}$$

of interarrival times. Suppose that each time a renewal occurs, a reward is received (e.g. for the occurrence of the renewal) or a cost is incurred (e.g. for replacement of the failed component). Denote the reward or cost associated with nth renewal by

$$Y_n,\ n = 1, 2, \ldots$$

Now Y_n will usually depend on X_n, the duration of the interarrival time between $(n-1)$th and nth renewal; but we assume that the pairs of random variables

$$\{X_n, Y_n\},\ n = 1, 2, \ldots$$

are independently and identically distributed. The sequence of the pairs of i.i.d. random variables

$$\{X_n, Y_n\},\ n = 1, 2, \ldots$$

is said to generate a *renewal reward process.*

Now
$$Y(t) = \sum_{n=1}^{N(t)} Y_n$$

gives the total reward earned (or cost incurred) by time t and

$$\{Y(t),\ t \geqslant 0\}$$

is called a renewal reward process generated by $\{X_n, Y_n\}$.

We have the following important theorem.

Theorem 6.11. Suppose that

$$E(X) = E(X_n) \text{ and } E(Y) = E(Y_n)$$

are finite. Then

$$\text{(a) with probability 1, } Y(t)/t \to E(Y)/E(X) \text{ as } t \to \infty$$

and

$$\text{(b) } EY(t)/t \to E(Y)/E(X) \text{ as } t \to \infty.$$

Proof of (a): We have

$$\frac{Y(t)}{t} = \frac{Y(t)}{N(t)} \cdot \frac{N(t)}{t}$$

$$= \frac{\sum_{n=1}^{N(t)} Y_n}{N(t)} \cdot \frac{N(t)}{t}.$$

By the strong law of large numbers, with probability 1,

$$\frac{\sum_{n=1}^{N(t)} Y_n}{N(t)} \to E(Y) \text{ as } t \to \infty.$$

and

$$N(t)/t \to 1/\mu = 1/E(X) \text{ as } t \to \infty \qquad \text{(by Theorem 6.4).}$$

Hence we have the result.

For proof of (b) reference may be made to some standard text, such as Ross (1970). ▲

Denote

$$\frac{E(Y)}{E(X)} = \xi$$

and

$$E(Y(t)) = C(t);$$

$C(t)$ is the average expected total reward (or cost) in $[0, t]$. Then from (b) above, we get, for large t,

$$C(t) = t\,\xi.$$

A more refined asymptotic form of $C(t)$ has recently been given by Christer (1978), who obtains the asymptotic expression for $C(t)$ in the form

$$C(t) = t\,\xi + \zeta,$$

where ξ and ζ are functions of the parameters and decision variables of the renewal process.

Example 6(a). *Age-based renewal reward process.* Consider a non-repairable item with lifetime X_i having distribution F. Suppose that replacements take place under an age replacement policy, with replacement interval T, and that the costs of replacements upon failure and replacements upon attainment of age T are A and B respectively, $(B < A)$. Then the associated age-based renewal reward process is given by $\{X_n, Y_n\}$, where

$$\begin{aligned} \Pr\{X_n \leqslant x\} &= F(x), \; x < T \\ &= 1 \qquad x \geqslant T \end{aligned}$$

and

$$\Pr\{Y_n = A\} = \Pr\{X_n < T\} = F(T)$$

$$\Pr\{Y_n = B\} = \Pr\{X_n \geqslant T\} = 1 - F(T).$$

We have

$$E(X) = E(X_n) = \int_0^T \{1 - F(x)\}\, dx$$

and

$$E(Y) = E(Y_n) = A\, F(T) + B\,\{1 - F(T)\}.$$

Then applying theorem 6.11, we have for large t,

$$\frac{C(t)}{t} \to \xi = \frac{E(Y)}{E(X)}.$$

In this case, the result

$$C(t)/t = \xi \; (t > 0),$$

can be obtained by straight forward calculation; the above provides an illustration of a reward renewal process. Christer (1978) gives an expression of the term ζ in the more refined asymptotic form

$$C(t) = t\,\xi + \zeta,$$

and this may be used to obtain better approximation in the asymptotic case.

6.7 Alternating (or Two-Stage) Renewal Process

The renewal processes so far considered are concerned with the occurrences of only one kind of event; the corresponding model relates to a system having only one state. For example, we considered the working of a component, the lifetime (or time to failure) being given by a sequence $\{X_n\}$ of i.i.d. random variables, on the assumption that the detection of failure and repair or replacement of the failed component take place *instantaneously*. Here the corresponding system has only one state—the working state and a renewal occurs at the termination of a working state (or failure of a component). Consider now that the detection and repair or replacement of a failed item are not instantaneous and that the time taken to do so is a random variable. The system then has two states—the working state and the repair state (during which repair of the failed component or search for a new one is under way). Here the two sequences of states—the working states and the repair (failed) states alternate. Suppose that the durations of the working states (or lifetimes or times to failure) are given by a sequence of i.i.d. random variables and the durations of repair states (times taken to repair or search) are given by a sequence of i.i.d. random variables. We have then an *alternating* or *two-stage* renewal process. As can be easily imagined, a great many renewal processes occurring in applications may be described by alternating renewal processes.

Consider an alternating renewal process. Suppose that the two alternating states are denoted by 1 and 0 and that the process starts at time $t = 0$ with state 1. Suppose that the duration of the two states 1 and 0 are given by the sequence of i.i.d. random variables X_n (having a common distribution F_1) and Y_n (having a common distribution F_0) respectively. Assume that a renewal of state i occurs at each termination of state i. Suppose that X's and Y's are independent. Denote by $N_i(t)$, $i = 1, 0$, the number of renewals of state i in $[0, t]$. Then $\{N_0(t), t \geqslant 0\}$ is an ordinary renewal process generated by the sequence of random variables $\{X_i + Y_i\}$ having distribution $H = F_1 * F_0$ (i.e. H is given by convolution of F_1 and F_0) and $\{N_1(t), t \geqslant 0\}$ is a modified renewal process with initial distribution F_1 (i.e. initial interarrival time X_1) and subsequent distribution $H = F_1 * F_0$ (i.e. subsequent interarrival times $X_{i+1} + Y_i$, $i = 1, 2, \ldots$).

Denote the renewal functions by

$$M_i(t) = E\{N_i(t)\}, \quad i = 1, 0.$$

Let f_1, f_0 be the p.d.f. of X_i and Y_i respectively and let $M_i^*(s)$, $f_i^*(s)$ be the Laplace transforms of the corresponding functions. We immediately get (using (4.2) and (1.5) respectively) that

$$M_1^*(s) = \frac{f_1^*(s)}{s\,\{1 - f_1^*(s)\, f_0^*(s)\}} \tag{7.1}$$

$$M_0^*(s) = \frac{f_1^*(s)\, f_0^*(s)}{s\,\{1 - f_1^*(s) f_0^*(s)\}}. \tag{7.2}$$

Theorem 6.12. For a system described by an alternating renewal process (starting with state 1 at $t = 0$) the probability that the system will be in state i at time t is given by

$$p_1(t) = M_0(t) - M_1(t) + 1 \tag{7.3}$$

and

$$p_0(t) = M_1(t) - M_0(t). \tag{7.4}$$

Proof: We first note that the rth renewal epoch of the renewal process $\{N_0(t), t \geqslant 0\}$ is given by $S_r = \sum_{i=1}^{r} (X_i + Y_i)$. The distribution of S_r is the r-fold convolution with itself of $H = F_1 * F_0$. Thus the distribution of S_r is given by H^{r*} and has p.d f. h_r (say).

Now, the event that the system is in state 1 at time t is equivalent to the following mutually exclusive events:

(A): the duration of the initial state 1 (in which the system started at $t = 0$) exceeds t, the probability of this event being given by

$$\Pr(A) = \Pr\{X_1 > t\} = 1 - F_1(t);$$

(B): in $(u, u + du)$, for some $u < t$, a renewal epoch S_r (for some $r = 1, 2, \ldots$) occurred and the system remained in state 1 for the remaining time $(t-u)$. The event (B) has the probability

$$\Pr(B)=\int_0^t \Big[\sum_{r=1}^{\infty}\{\Pr(u\leqslant S_r<u+du)\}\Big]\Pr(X_1>t-u)].$$

The probability $p_1(t)$ of the event that the system is in state 1 at time t is given by

$$p_1(t)=1-F_1(t)+\int_0^t \sum_{r=1}^{\infty} h_r(u)\{1-F_1(t-u)\}\,du. \tag{7.5}$$

Since L.T. of $h_r(u)$ is $h_r^*(s)=[f_1^*(s)f_0^*(s)]^r$, we have, taking Laplace transforms of (7.5),

$$\begin{aligned}p_1^*(s)&=\left\{\frac{1}{s}-\frac{f_1^*(s)}{s}\right\}+\sum_{r=1}^{\infty}[f_1^*(s)f_0^*(s)]^r\left[\frac{1}{s}-\frac{f_1^*(s)}{s}\right]\\&=\frac{1-f_1^*(s)}{s}\left[1+\frac{f_1^*(s)f_0^*(s)}{1-f_1^*(s)f_0^*(s)}\right]\\&=\frac{1-f_1^*(s)}{s\{1-f_1^*(s)f_0^*(s)\}}.\end{aligned} \tag{7.6}$$

From (7.1) and (7.2), we find that the above can also be expressed as

$$p_1^*(s)=M_0^*(s)-M_1^*(s)+\frac{1}{s} \tag{7.7}$$

from which we get, on inversion,

$$p_1(t)=M_0(t)-M_1(t)+1.$$

It follows that

$$\begin{aligned}p_0(t)&=1-p_1(t)\\&=M_1(t)-M_0(t).\end{aligned}$$

The theorem is thus proved. ▲

Corollary. *Steady State Distribution; Limiting Value of* $p_i(t)$.
If $E(X)$, $E(Y)$ exist finitely, then

$$\lim_{t\to\infty} P_1(t)=\frac{E(X)}{E(X)+E(Y)}$$

$$\lim_{t\to\infty} P_0(t)=\frac{E(Y)}{E(X)+E(Y)}.$$

Proof: Let $E(X)=a$, $E(Y)=b$. Then

$$f_1^*(s)=1-as+o(s)$$

$$f_0^*(s)=1-bs+o(s).$$

Substituting in (7.6), we get

$$p_1{}^*(s) = \frac{as + o(s)}{s\,\{1 - 1 + (a + b)\,s + o(s)\}}$$

$$= \frac{1}{s} \cdot \frac{a}{a+b} + o\left(\frac{1}{s}\right).$$

Now
$$\lim_{t\to\infty} p_1(t) = \lim_{s\to 0} sp_1{}^*(s)$$

$$= \frac{a}{a+b}$$

and thus

$$\lim_{t\to\infty} p_1(t) = \frac{E(X)}{E(X) + E(Y)}. \tag{7.8}$$

It follows that

$$\lim_{t\to\infty} p_0(t) = \frac{E(Y)}{E(X) + E(Y)}. \tag{7.9}$$

Note: Further generalisations of this alternating renewal process have led to interesting results. One such generalisation is that the system has N states $1, 2, \ldots i, \ldots, j, \ldots, N$, such that transition from one state i to any other state j takes place with transition probability

$$p_{ij}, \quad \sum_{j=1}^{N} p_{ij} = 1,$$

and that the time spent in state i before moving to state j is a random variable whose distribution depends on both these states. We have then a semi-Markov process.

Markov renewal process is another important generalisation of the renewal process. We deal with these two topics in a separate chapter.

6.8 Regenerative Stochastic Processes: Existence of Limits

In the Corollary to Theorem 6.12, we were concerned with finding steady-state behaviour or limiting distribution of $p_i(t)$, as $t \to \infty$, $i = 0, 1$. Steady state distribution for stochastic processes in the more general case is of considerable importance. Even without finding the exact limiting distribution, it is helpful to know about the existence of the limits in stochastic processes. Renewal theorem provides a very simple and powerful tool in proving the existence of limits or of steady state for a large class of stochastic processes. The result is given below as an existence theorem.

Definition. Let $\{X(t)\}$ be a stochastic process having a countably infinite number of states $\{0, 1, 2, \ldots, k, \ldots\}$. Assume that there are epochs $t_1, t_2, \ldots$ at which the process probabilistically restarts from scratch, i.e. with probability one there exists an epoch t_1 such that the process beyond t_1 is a probabilistic replica of the whole process starting at $t = 0$; then the same property holds for $t_2, t_3, \ldots$. Such a process is known as a *regenerative process*. Clearly $\{t_n\}$ forms a renewal process and every time a renewal occurs a *cycle* is said to be completed.

Theorem 6.13. Let $\{X(t)\}$ be a regenerative process and let

$$p_k(t) = \Pr\{X(t) = k\}.$$

Then $\lim\limits_{t\to\infty} p_k(t) = p_k$ exists such that

$$p_k \geqslant 0 \text{ and } \sum_{k=0}^{\infty} p_k = 1.$$

Proof: Let $q_k{}'t) = \Pr\{X(t) = k,\ t_1 > t\}$. Then

$$\sum_{k=0}^{\infty} q_k(t) = \Pr\{t_1 > t\} = 1 - F(t),$$

where $F(\cdot)$ is the d.f. of the interarrival time

$$(t_n - t_{n-1}) = T_n$$

having mean, say, μ. We have, by conditioning on t_1,

$$p_k(t) = \int_0^\infty \Pr\{X(t) = k \mid t_1 = s\}\, dF(s)$$

$$= q_k(t) + \int_0^t p_k(t - x)\, dF(x); \tag{8.1}$$

since

$$q_k(t) = \Pr\{X(t) = k,\ t_1 > t\}$$

$$= \int_0^\infty \Pr\{X(t) = k, t_1 > t \mid t_1 = s\}\, dF(s)$$

$$= \int_t^\infty \Pr\{X(t) = k \mid t_1 = s\}\, dF(s).$$

From Theorem 6.3. we see that the renewal equation (8.1) has a unique solution

$$p_k(t) = q_k(t) + \int_0^t q_k(t - x)\, dM(x), \tag{8.2}$$

where $M(x)$ is the renewal function of the renewal process with distribution F. Now $q_k(t)$ is non-negative, non-increasing and tends to 0 as $t \to \infty$, and $\int_0^\infty q_k(t)\, dt < \infty$. Thus application of the key renewal theorem (Theorem 6.7) to (8.2) at once yields

$$\lim_{t\to\infty} p_k(t) \to \frac{1}{\mu}\int_0^\infty q_k(x)\,dx, \tag{8.3}$$

i.e. $\lim_{t\to\infty} p_k(t) = p_k$ exists. Clearly, $p_k \geqslant 0$,

and $$\sum_{k=0}^{\infty} p_k = \frac{1}{\mu}\int_0^\infty \{\sum_{k=0}^{\infty} q_k(x)\}\,dx = \frac{1}{\mu}\int_0^\infty \{1-F(x)\}\,dx = 1.$$

Hence proved. ▲

Note: Clearly, μ is the expected duration of a complete cycle, and

$\int_0^\infty q_k(t)\,dt = \int_0^\infty \Pr\{X(t)=k,\, t_1 > t\}\,dt =$ expected duration of sojourn (or visit to) state k during a cycle $= E(\tau_k)$, say. Thus

$$p_k = \lim_{t\to\infty} \Pr\{X(t)=k\} = \frac{E(\tau_k)}{\mu}. \tag{8.4}$$

Example 8 (a). *Queueing process** with arrival and service rates λ and μ and $\lambda/\mu < 1$. Let the state of the queueing process at epoch t be given by the number in the system $X(t)$ at that epoch; then the states of the process are $i \in E = \{0, 1, 2, \ldots\}$. Suppose that time is reckoned from the epoch 0 at which an arriving customer finds the system empty with a free server and the service begins, i e. with commencement of a busy period (with distribution, say, F_1). The busy period B is followed by an idle period I during which no customer is present in the system (I having distribution. say, F_0). A busy period and an idle period alternate and together constitute a *busy cycle* the whole—process restarts itself (probabilistically) from scratch at the end of each busy cycle. An idle period corresponds to the visit to state 0 while a busy period to states $i \neq 0$. Clearly, $q_i(t) = 1 - F_i(t)$,

$$\int_0^\infty q_0(t)\,dt = \text{Expectation of the duration of idle period} = E(I),$$

and $$\int_0^\infty q_1(t)\,dt = \text{Expectation of the duration of busy period} = E(B).$$

Thus for such a system

$$p_0 = \lim_{t\to\infty} \Pr\{X(t)=0\} = \frac{E(I)}{E(I)+E(B)}$$

which is a very interesting result.

---

*See Sec. 10.1 for description.

In this example we were concerned with practically an alternating (two-stage) system; however the theorem is applicable to multi-stage systems also.

Generalisation of the classical renewal theory: The classical theory (to which we devoted our attention) deals with a sequence of *i.i.d. non-negative* r.v.'s $\{X_i\}$ and considers the asymptotic behaviour of $E\{N(t)\}$ as $t \to \infty$ (where $N(t) = \sup\{n: S_n \leqslant t\}$, $S_n = \sum_{i=1}^{n} X_i$, $E\{N(t)\} = \sum_{n=0}^{\infty} \Pr\{S_n \leqslant t\}$). Feller (1941) obtained the elementary renewal theorem which states that as $t \to \infty$, $E\{N(t)\}/t \to 1/\mu$, where $\mu = E(X_i)$. This result of classical theory was generalised step by step in several directions. We refer here to some of the works of the several researchers who made very significant contribution of immense mathematical and theoretical interest. This has been done to indicate the broad areas and to draw attention to some of the works for the benefit of those who might be interested in pursuing the topic more deeply. First generalisation was concerned with sequence of r.v.'s $\{X_i\}$ which were *not* assumed to be non-negative, i.e. generalisation from distributions concentrated on the half-line to distributions on the whole line. Feller considers what he terms 'transient distribution' F on the whole line.

Important generalisation of renewal theory to various cases of transient distributions, are due, among others to Chung *et al.* (*Ann. Math.* **55** (1952) 1–6, *Proc. Am. Math. Soc.* **3** (1952) 303–309), Blackwell (*Pacific J. Math.* **3** (1953) 315–320), Feller *et al.* (*J. Math. Mech.* **10** (1961) 619–624). We state below (without proof) an important *extension of Blackwell's renewal theorem.*

Let $\{X_i\}$ be a sequence of (continuous) i.i.d. random variables which assume values in the range $(-\infty < x < \infty)$ such that $0 < \mu = E(X_i) \leqslant \infty$. Let $U(t) = \sum_{n=0}^{\infty} \Pr\{S_n \leqslant t\}$, $-\infty < t < \infty$, where $S_n = \sum_{i=1}^{n} X_i$. Then for $h > 0$,

$$U(t+h) - U(t) = \frac{h}{\mu} \quad \text{as} \quad t \to \infty$$

$$U(t+h) - U(t) = 0 \quad \text{as} \quad t \to -\infty.$$

It is to be noted that in this case $U(t)$ is, in general, not equal to $E\{N(t)\}$, where $N(t) = \sup\{n: S_n \leqslant t\}$ (for details of proof reference may also be made to the books by Prabhu and Feller Vol. II).

Attention was also turned towards the nature of $\{X_i\}$. Chow and Robbins (*Ann. Math. Statis.* **34** (1963) 390–395) investigated the conditions on the joint distribution of X_i which would ensure the validity of the elementary renewal theorem,

$$\sum_{n=0}^{\infty} \Pr\{S_n \leqslant t\}/t \to 1/\mu$$

for some positive constant μ. They relaxed the restriction of independence of the r.v.'s X_i and obtained a renewal theorem for the dependent case as well as for a sequence of nonidentical r.v.'s. Smith (*Pacific J. Math.* **14** (1964) 673–699) also extended the theorem to non-identically distributed r.v.'s. Heyde (*Ann. Math. Statis.* **37** (1966) 699–710) considered asymptotic behaviour of more general forms like

$$\sum_{n=0}^{\infty} a_n \Pr\{S_n \leqslant t\} \quad \text{and} \quad \sum_{n=0}^{\infty} a_n \Pr\{\max_{1 \leqslant k \leqslant n} S_k \leqslant t\}$$

for a sequence of i.i.d. r.v.'s $\{X_i\}$ and for a sequence of positive constants $\{a_n\}$ such that Σa_n diverges.

Later in a series of papers Sankaranarayanan *et al.* (*Pacific J. Math.* **30** (1969) 785–803, *Act. Math. Aca. Sc. Hung.* **31** (1978) 1–8, *Math. Oper. Statist. Ser. Opt.* **8** (1977) 581–590) considered suitable sequences of $\{a_n\}$ and studied the asymptotic behaviour of such functions for sequences of correlated and dependent r.v.'s, as well as for stationary discrete stochastic processes. Renewal theory for Markov chains has been investigated by Kesten (*Ann. Prob.* **3** (1974) 355–387) and Athreya *et al.* (*Ann. Prob.* **6** (1978) 788–797) etc. Ryan Jr. (*Ann. Prob.* **4** (1976) 656–661) has considered the multidimensional case, while Chow *et al.* (*Ann. Prob.* **7** (1979) 304–318) have developed an extended renewal theory. Thus extensions have been widely varied.

The above is an illustrative (and by no means an exhaustive) account of the many interesting and far-reaching contributions made by several workers towards generalisation and extension of the renewal theory.

EXERCISES

6.1 Show that

$$\Pr\{N_e(t) = r\} = \frac{1}{\mu}\int_0^T [\Pr\{N(u) = r-1\} - \Pr\{N(u) = r\}]\,du \qquad r = 1, 2, 3, \ldots$$

and $$\Pr\{N_e(t) = 0\} = 1 - \frac{1}{\mu}\int_0^T \Pr\{N(u) = 0\}\,du.$$

6.2 Taking $f(t)$, $g(t)$ as given in Example 4(a), obtain $M_D(t)$ by using the relation (4.3).

6.3 Show that, the renewal function of a renewal process generated by X_n, $n = 1, 2, \ldots$ is given by

$$M(t) = \frac{t}{\mu} + \frac{\sigma^2 - \mu^2}{2\mu^2} + o(1),$$

where $\mu\ (< \infty)$ and σ^2 are the mean and the variance of X_n. Further, show that $\operatorname{var}\{N(t)\} = (\sigma^2/\mu^3)\,t + o(t)$.

6.4 Find the second moment $L(t) = E\{(N(t))^2\}$ for the renewal process with mixed expotential distribution having p.d.f.

$$f(t) = pae^{-at} + (1-p)be^{-bt},\ 0 \leqslant p \leqslant 1,\ a > b > 0.$$

6.5 Put forward direct probabilistic argument to show that, in the Markovian case,

$$\Pr\{N(t) \geqslant n\} = \Pr\{N_A(t, T) \geqslant n\} = \Pr\{N_B(t, T) \geqslant n\}$$

for all $t \geqslant 0$, $T \geqslant 0$, $n = 0, 1, 2, \ldots$.

6.6 Show that, for $t \geqslant 0$, $T \geqslant 0$, $n = 0, 1, 2, \ldots$ both (i) and (ii)

(i) $\Pr\{N(t) > n\} \geqslant \Pr\{N_A(t, T) > n\}$

(ii) $\Pr\{N(t) > n\} \geqslant \Pr\{N_B(t, T) > n\}$

hold *iff* F is NBU, i.e. the lifetime X of the unit satisfies

$\Pr\{X > x + y\} \leqslant \Pr\{X > x\} \Pr\{Y > y\}$ for $x \geqslant 0, y \geqslant 0$

(Barlow and Proschan, 1975).

6.7 Let $R_A(t, T)$, $R_B(t, T)$ be the number of removals in $[0, t]$ under age and block replacement policies respectively. Show that

$$\Pr\{N(t) \geqslant n\} \leqslant \Pr\{R_A(t, T) \geqslant n\} \leqslant \Pr\{R_B(t, T) \geqslant n\}$$

for $t \geqslant 0$, $T \geqslant 0$, $n = 0, 1, 2, \ldots$.

(The intuitive meaning of the above is quite clear.)

(Barlow and Proschan, 1964).

6.8 Show that

$$\lim_{t\to\infty} \frac{E\{R_A(t, T)\}}{t} = \frac{1}{\int_0^T \{1 - F(x)\}\, dx},$$

$$\lim_{t\to\infty} \frac{E\{R_B(t, T)\}}{t} = \frac{M(t) + 1}{T}$$

and that for all $T > 0$,

$$M(T) \geqslant \frac{T}{\int_0^T \{1 - F(x)\}\, dx} - 1$$

$$\geqslant T/\mu - 1$$ (Barlow and Proschan, 1964).

6.9 Consider the number N of renewals in random time T having gamma distribution and p.d.f.

$$q(t) = \frac{a^k t^{k-1} e^{-at}}{\Gamma(k)},\ k = 1, 2, \ldots$$

Show that the p.g.f. of N is given by

$$G(z) = \frac{a^k}{\Gamma(k)} \left(-\frac{\partial}{\partial s}\right)^{k-1} \left[\frac{1 - f^*(s)}{s\{1 - zf^*(s)\}}\right]_{s=a},$$

Examine the truth of the following assertion:
N has negative binomial distribution with p.g.f.

$$G(z) = \left[\frac{1-f^*(a)}{1-zf^*(a)}\right]^k$$

if and only if the renewal process $\{X_i\}$ is Poisson (i.e. X_i has exponential distribution with p.d.f., say $f(t) = ce^{-ct}$).

6.10 Distribution of residual lifetime $Y(t)$: alternative form.
Show that

$$\Pr\{Y(t) \leqslant x\} = \int_t^{t+x} \{1 - F(t+x-y)\}\, dM(y).$$

6.11 If $Z(t)$ is the spent lifetime at time t for the process $\{X_i\}$ and $\mu = E(X) < \infty$, then, with probability 1

$$\frac{Z(t)}{t} \to 0 \quad \text{as} \quad t \to \infty$$

and that

$$E\left[\frac{Z(t)}{t}\right] \to 0 \quad \text{as} \quad t \to \infty.$$

6.12 Show that for the renewal reward process $\{X_n, Y_n\}$, the asymptotic form of $C(t)$ is given by

$$C(t) = t\xi + \zeta,$$

where $\xi = E(Y)/E(X)$

and $$\zeta = \frac{E(Y)\text{--}E(X^2)}{2\{E(X)\}^2} - \lim_{t\to\infty} E\{Y_{N(t)+1}\},$$

and that for the age-based renewal reward process considered in Example 6(a)

$$\zeta = -\left[\frac{2TE(X) - E(X^2)\, E(Y)}{2\{E(X)\}^2} - \frac{A\{T - E(X)\}}{E(X)}\right]$$

(Christer, 1978).

6.13 Find $M_i(t)$, $i = 1, 0$ for the alternating renewal process $\{X_n, Y_n\}$ with distribution of X_n and Y_n given by

$$F_1(x) = 1 - e^{-ax}, \text{ and } F_0(x) = 1 - e^{-bx} \text{ respectively.}$$

Show that

$$p_1(t) = \frac{b}{a+b} + \frac{a}{a+b}\, e^{-(a+b)t}$$

$$p_0(t) = 1 - p_1(t) = \frac{a}{a+b} - \frac{a}{a+b}\, e^{-(a+b)t}.$$

Further that

$$\lim_{t\to\infty} p_1(t) = \frac{b}{a+b} = \frac{E(X)}{E(X)+E(Y)}$$

$$\lim_{t\to\infty} p_0(t) = \frac{a}{a+b} = \frac{E(Y)}{E(X)+E(Y)}.$$

S.C.E.R.T., W
Library
Calcutta
25/3, B. C. Roa

Note that $p_1(t)$ for the Markovian case was obtained earlier by using results of Markov processes (see Example 5(b), Ch. 4), and that the result giving the limiting value of $p_1(t)$ as $t \to \infty$ holds independently of the form of the distributions of X and Y (see Corollary to Theorem 6.12).

6.14 Show that the number of renewals $N(t)$ in $[0, t]$ of a renewal process induced by X_i is asymptotically normal with mean t/μ and variance $(\sigma^2/\mu^3)\, t$, μ and σ^2 being the mean and the variance of X_i.

6.15 For block replacement (with replacement interval T) with costs of replacement upon failure and at intervals $T, 2T, \ldots$ being A and B respectively, find $E\{Y(t)\}/t$ as $t \to \infty$, where $Y(t)$ is the total cost incurred by time $t \geqslant 0$.

6.16 Prove that the joint distribution of the residual lifetime $Y(t)$ and the spent lifetime $Z(t)$, is given by

$$\Pr\{Z(t) > x,\ Y(t) > y\} \to \frac{1}{\mu} \int_{x+y}^{\infty} \{1 - F(u)\}\, du \qquad \text{as } t \to \infty \ (x > 0,\ y > 0)$$

(where F denotes the distribution of the lifetime with mean μ).

REFERENCES

A.A. Balkema and L. De Haan, "Residual lifetime at great age," *Ann. Prob.*, **2**, 792-804 (1974).

R.E. Barlow and F. Proschan, "Comparison of replacement policies and renewal theory implications," *Ann. Math. Stat.*, **35**, 577–589 (1964).

———*Statistical Theory of Reliability and Life Testing*, Holt, Rinehart & Winston, New York (1975).

A.H. Christer, "Refined asymptotic costs for renewal reward processes," *J. Opns. Res. Soc.*, **29**, 577–583 (1978).

D.R. Cox, *Renewal Theory*, Methuen, London (1962).

W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. II, Wiley, New York (1966).

U.S. Karmarkar, "Future costs of service contracts for consumer durable goods," *AIIE Trans.*, **10**, 380–387 (1978).

N.U. Prabhu, *Stochastic Processes*, Macmillan, New York (1965).

S.M. Ross, *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, (1970).

A.N. Sirjaev, *Statistical Sequential Analysis*, American Mathematical Society, Providence, R.I. (1973).

W.L. Smith, "Renewal Theory and its Ramifications," *J.R.S.S-B.* **20**, 243–302 (1958).

CHAPTER 7

MARKOV RENEWAL AND SEMI-MARKOV PROCESSES

7.1 Introduction

We shall consider here some kind of generalisation of a Markov process, as well as of a renewal process. We first consider a Markov process with discrete state space, and its generalisation by way of random transformation of the time scale of the process. Let $X(t)$ be a Markov process with discrete state space and let the transitions occur at epochs (or instants of time) t_i, $i = 0, 1, 2, \ldots$. Then the sequence $\{X_n = X(t_n + \ )\}$ forms a Markov chain and the time intervals $(t_{n+1} - t_n) = \tau_n$ between transitions have exponential distributions, the parameter of which may depend on X_n. We shall now suppose that the transitions $\{X_n\}$ of the process $X(t)$ still constitute a Markov chain, but that the transition intervals $(t_{n+1} - t_n)$ have any distribution, with parameter which may depend not only on X_n but also on X_{n+1}. The process $\{X_n, t_n\}$ (as well as $\{X_n, \tau_n\}$) is then called a *Markov renewal process* and the process $X(t)$ is known as a *semi-Markov process*. A semi-Markov process (s-M.P.) is thus a stochastic process in which changes of state occur according to a Markov chain and in which the time interval between two successive transitions is a random variable, whose distribution may depend on the state from which the transition takes place as well as on the state to which the next transition takes place. The s-M.P. provides a convenient way of describing the underlying Markov renewal process. Semi-Markov processes were first studied by Lévy and Smith, while Markov renewal processes by Pyke. Considerable interest in these topics has grown over the years, as, such generalisation of Markov process can be fruitfully applied in building models of many phenomena. Many problems in queueing, reliability and inventory theories may be approached through Markov renewal and semi-Markov processes (see Cinlar 1969, 1975 for a survey). In stochastic models for many types of social behaviour these processes have been used (see Bartholomew, 1967). For example, problems of occupational mobility, where transitions between occupations as well as intervals between transitions are taken into account, may be treated by means of semi-Markov processes. So also in some studies of problems relating to market behaviour. A semi-Markov model has been used by Kao (1974) in some context of hospital administration—in the study of the dynamics of movement of patients through various care zones (e.g. medical unit, surgical unit, intensive care unit etc.) in a hospital. The theory of birth has been developed as a Markov renewal process by Sheps (1964) (also Sheps *et al.*, 1973), and the theory has been applied to microdemography by Potter (1970) (see also Keyfitz, 1977). In fact, the importance of semi-Markov processes, in particular, and Markov renewal theory, in general, lies to a large extent in the wide domain of applications rather than in the richness of their theoretical developments.

7.2 Definitions and Preliminary Results

Definitions. Let the states of a process be denoted by the set $E = \{0, 1, 2, \ldots\}$, and let the transitions of the process occur at epochs (or instants of times) $t_0 = 0, t_1, t_2, \ldots (t_n < t_n + 1)$. Let X_n denote the transition occurring at epoch t_n. If

$$\Pr\{X_{n+1} = k,\ t_{n+1} - t_n \leqslant t \mid X_0 = .,\ \ldots, X_n = .\,;\ t_0, \ldots, t_n\}$$
$$= \Pr\{X_{n+1} = k,\ t_{n+1} - t_n \leqslant t \mid X_n = .\}, \tag{2.1}$$

then $\{X_n, t_n\}$, $n = 0, 1, 2, \ldots$ is said to constitute a *Markov renewal process* with state space E. $\{X_n, \tau_n\}$ is also used to denote the process $\{X_n, t_n\}$. Assume that the process is temporally (or time) homogeneous, that is,

$$\Pr\{X_{n+1} = k,\ t_{n+1} - t_n \leqslant t \mid X_n = j\} = Q_{jk}(t) \tag{2.2}$$

is independent of n. Define

$$p_{jk} = \lim_{t\to\infty} Q_{jk}(t) = \Pr\{X_{n+1} = k \mid X_n = j\}. \tag{2.3}$$

(Symbols $Q_{j,k}$ for Q_{jk} and $p_{j,k}$ for p_{jk} will also be used.)

Proposition (1):
$\{X_n, n = 0, 1, 2, \ldots\}$ constitutes a Markov chain with state space E and t.p.m. $P = (p_{jk})$.

The continuous parameter process $Y(t)$ with state space E, defined by

$$Y(t) = X_n \quad \text{on} \quad t_n \leqslant t < t_{n+1} \tag{2.4}$$

is called a *semi-Markov process*. The Markov chain $\{X_n\}$ is said to be an *embedded Markov chain* of the semi-Markov process. X_n refers to the state of the process at transition occurring at epoch t_n and $Y(t)$ that of the process at its most recent transition. Consider, for example, a simple random walk performed by a particle. Suppose that $t_0, t_1, \ldots,$ are the epochs when transitions take place and that X_n is the transition at the epoch t_n; also that at t_n the particle makes a move to the right (and enters the state j), at t_{n+1} to the left (enters the state $j-1$), at t_{n+2} to the right (enters the state j), at t_{n+3} to the right (enters the state $j+1$), at t_{n+4} moves to the right (enters the state $j+2$), and so on.

Then
$$\begin{aligned} Y(t) &= j, & t_n &\leqslant t < t_{n+1}\\ &= j-1, & t_{n+1} &\leqslant t < t_{n+2}\\ &= j, & t_{n+2} &\leqslant t < t_{n+3}\\ &= j+1, & t_{n+3} &\leqslant t < t_{n+4}\\ &= j+2, & t_{n+4} &\leqslant t < t_{n+5}\end{aligned}$$

and so on. The position of the particle is shown in Fig. 7.1.

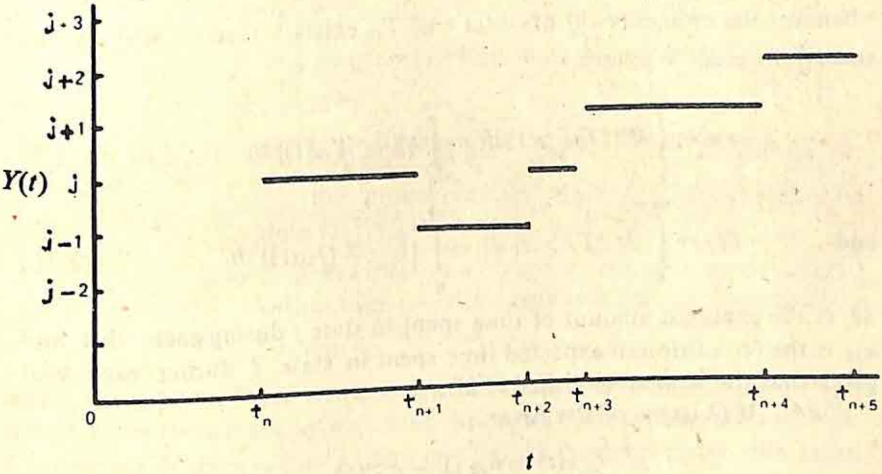


Fig. 7.1

Example 2(a). A pure birth process is a Markov renewal process with
$$\begin{aligned} Q_{jk}(t) &= 1 - e^{-a_j t},\ k = j+1\\ &= 0, \text{ otherwise.}\end{aligned}$$

Unless otherwise stated, it will be assumed in the following sections, that the state space of the process is finite having m states, $1, 2, \ldots, m$.

Waiting Times

Let the time spent by the system in state j before its next transition, given that the next transition is into state k, be a random variable T_{jk} having distribution function (d.f.):

$$W_{jk}(t) = \Pr\{T_{jk} \leqslant t\} = \Pr\{t_{n+1} - t_n \leqslant t \mid X_n = j, X_{n+1} = k\},\quad j, k = 1, 2, \ldots, m. \tag{2.5}$$

The random variable T_{jk} (called *sojourn time* or *waiting time*) depends on the state X_n being visited and the state X_{n+1} to be entered in the very next transition. Assume that if $p_{jk} = 0$ for some j, k; then $Q_{jk}(t) = 0$ for all t and $W_{jk}(t) = 1$. Then

$$W_{jk}(t) = Q_{jk}(t)/p_{jk}. \tag{2.6}$$

The variable T_{jk} denotes the conditional waiting time in state j given that the next state is k; and if T_j denotes the unconditional waiting time in state j, then

$$T_j = \sum_k p_{jk} T_{jk} \tag{2.7}$$

and $$W_j(t) = \Pr\{T_j \leqslant t\} = \sum_k p_{jk} W_{jk}(t) = \sum_k Q_{jk}(t). \tag{2.8}$$

We also have

$$M_j^{(r)} = E(T_j^r) = \sum_k p_{jk} E(T_{jk}^r) = \sum_k p_{jk} \alpha_{jk}^{(r)},$$
$$r = 1, 2, \ldots$$

whenever the moments $\alpha_{jk}^{(r)}$ of order r of T_{jk} exists. Denote $M_j^{(1)}$ by M_j and $\alpha_{jk}^{(1)}$ by α_{jk}. We have

$$\alpha_{jk} = \int_0^\infty \Pr\{T_{jk} > t\}\, dt = \int_0^\infty \{1 - W_{jk}(t)\}\, dt \tag{2.9}$$

and $$M_j = \int_0^\infty \Pr\{T_j > t\}\, dt = \int_0^\infty \{1 - \sum_k Q_{jk}(t)\}\, dt; \tag{2.10}$$

M_j is the expected amount of time spent in state j during each visit and α_{jk} is the (conditional) expected time spent in state j during each visit given that the next state visited is k.

Note: If $Q_{jk}(t)$ is of the form

$$Q_{jk}(t) = p_{jk}\{1 - e^{-a_j t}\}$$

(equivalently if $W_{jk}(t)$ has the form of the second factor), then it can be shown that

$$\Pr\{Y(t+s) = k \mid Y(u), u \leqslant t\} = \Pr\{Y(t+s) = k \mid Y(t)\}$$

and that

$$\Pr\{Y(t+s) = k \mid Y(t) = j\} = p_{jk}(s)$$

is independent of t. In this case the semi-Markov process becomes a time-homogeneous Markov process with t.p.m. $P = (p_{jk})$. (In fact, this is what we get in Example 2(b) given below.) The Markov property which holds, in case of semi-Markov processes only at the epochs t_n, will hold for all t *only when* $Q_{jk}(t)$ is of the above form. Thus in semi-Markov processes there is freedom in the distribution of the sojourn times T_{jk}, but only at the expense of the Markov property, which instead of holding for all t holds only at the jump epochs t_n.

A Markov process is a semi-Markov process but the converse is not true.

Suppose that $X(t)$ is a s-M.P. with discrete time, and that the transition intervals $(t_{n+1} - t_n)$ are equal to unity for all n. Then the sequence $\{X_n\}$, where $X_n = X(t_n +\ \ )$ reduces to a simple Markov chain.

Example 2(b). The $M/M/1$ queue*. Suppose that customers arrive at a counter in accordance with a Poisson process with parameter λ and are served by a single server, the service times of customers being i.i.d. exponential random variables with mean $1/\mu$. Then $\{N(t), t > 0\}$, where $N(t)$ is the number of customers in the system at time t, is a s-M.P. (it is also a Markov process). The states of the s-M.P. are $0, 1, 2, \ldots$. Suppose that $t_0, t_1, \ldots$ are the successive instants at which changes of states (by way of arrival or departure of a unit) take place and

---

*For description, see Ch. 10.

$$\begin{aligned}
X_n &= N(t_n + 0);\\
Q_{0,1}(t) &= \Pr\{X_{n+1} = 1, t_{n+1} - t_n \leqslant t \mid X_n = 0\}\\
&= \Pr\{\text{the process enters state 1 from state 0 by time } t\}\\
&= 1 - e^{-\lambda t},\\
Q_{j,j-1}(t) &= \Pr\{X_{n+1} = j-1, t_{n+1} - t_n \leqslant t \mid X_n = j\}, j \geqslant 1\\
&= \Pr\{\text{the process enters state } j-1 \text{ from staie } j \text{ by time } t\}\\
&= \Pr\{\text{one transition (service completion or arrival) within time } t\} \cdot \Pr\{\text{transition is service completion}\}\\
&= \Pr\{\min(A, B) \leqslant t\} \cdot \Pr\{A > B\},
\end{aligned}$$

where the random variables A, B denote respectively the interarrival and service times (which are independent exponential with mean $1/\lambda$ and $1/\mu$). Since $\Pr\{A > B\} = \mu/(\lambda + \mu)$ and $\min(A, B)$ is exponential with mean $1/(\lambda + \mu)$, we have

$$Q_{j,j-1}(t) = \{1 - e^{-(\lambda+\mu)t}\} \cdot \{\mu/(\lambda + \mu)\}, j > 1.$$

Similarly

$$\begin{aligned}
Q_{j,j+1}(t) &= \Pr\{\min(A, B) \leqslant t\} \cdot \Pr\{B > A\}\\
&= \{1 - e^{-(\lambda+\mu)t}\} \cdot \{\lambda/(\lambda + \mu)\}, j > 1
\end{aligned}$$

Further $\quad Q_{jk}(t) = 0$, for $k \neq j-1, j+1\ (j, k > 1)$.

Hence the transition probabilities of the embedded Markov chain $\{X_n\}$ are

$$\begin{aligned}
p_{0,1} &= 1, p_{jk} = 0, k \neq j-1, j+1\\
p_{j,j-1} &= \mu/(\lambda + \mu), p_{j,j+1} = \lambda/(\lambda + \mu), j \geqslant 1.
\end{aligned}$$

Thus

$$\begin{aligned}
W_{0,1}(t) &= \Pr\{t_{n+1} - t_n \leqslant t \mid X_n = 0, X_{n+1} = 1\}\\
&= 1 - e^{-\lambda t}.\\
W_{j,j-1}(t) &= 1 - e^{-(\lambda+\mu)t} = W_{j,j+1}(t), j \geqslant 1.
\end{aligned}$$

Further, T_0, the unconditional waiting time in state 0 is the same as the interarrival time, i e. T_0 is exponential with mean $1/\lambda$. T_j $(j \geqslant 1)$, the unconditional waiting time in state j is the time required for a transition (either an arrival or service completion). Thus T_j which is equivalent to $\min(A, B)$, is exponential with mean $1/(\lambda + \mu)$. We have

$$\begin{aligned}
M_0 &= E(T_0) = 1/\lambda\\
M_j &= E(T_j) = 1/(\lambda + \mu), j \geqslant 1.
\end{aligned}$$

Example 2(c). The $M/G/1$ queue. Here we suppose that the service time distribution is general, other things being the same as in the $M/M/1$ queue. We note that $\{N(t), t \geqslant 0\}$ is *not* a semi-Markov process, because the state of the system after a transition depends not only on the state of $N(t)$, but also on the amount of elapsed service time of the person receiving service, if any. If we consider that transitions take place only when a customer completes his service, there will be no elapsed service time. X_n will here give the number of customers in the system at the instant that nth customer departs. If $Y(t)$ denotes the number of customers left-behind in the system by the most recent departure, $Y(t) = X_n$, $t_n \leqslant t < t_{n+1}$ and $Y(t)$ will be a semi-Markov process having X_n, $n = 0, 1, \ldots$ for its embedded Markov chain. $\{X_n, t_n\}$, $n = 0, 1, 2, \ldots$ is a Markov renewal process. The sequence of intervals $(t_{n+1} - t_n)$, $n = 0, 1, 2, \ldots$ being the service times of successive units (or equivalently $\{t_n\}$, $n = 0, 1, 2, \ldots$) forms a renewal process (as explained below).

We now examine how Markov renewal processes can be viewed as a generalisation of renewal process. It can be shown that, for $n = 1, 2, \ldots$

$$\begin{aligned} \Pr\{t_1 - t_0 \leqslant u_1, t_2 - t_1 \leqslant u_2, \ldots, t_n - t_{n-1} \leqslant u_n \\ X_0 = i_1, X_1 = i_2, \ldots, X_n = i_n\} \\ = W_{i_1 i_2}(u_1)\, W_{i_2 i_3}(u_2) \ldots W_{i_{n-1} i_n}(u_n). \end{aligned} \tag{2.11}$$

In particular, if there is only one state, then the transition intervals $(t_n - t_{n-1})$ are i.i.d. non-negative random variables.

Proposition (2):

If E consists of a single point, then the sequence $\{t_n\}$, $n = 0, 1, 2, \ldots$ constitutes a renewal process. Suppose $j \in E$ is fixed and $S_1^j, S_2^j, \ldots$ are the successive epochs t_n for which $X_n = j$. Then for each fixed j. the sequence $\{S_n^j\}$, $n = 0, 1, 2, \ldots$ constitutes a renewal process which is ordinary if j is the initial state (with $S_0{}^j = T_0 = 0$) and delayed, otherwise.

Proposition (3):

To each state $j \in E$, there corresponds (a possibly delayed) renewal process $\{S_n^j\}$, $j = 1, 2, \ldots, m$ and the superposition of all these renewal processes gives the set of transition epochs $t_0, t_1, t_2, \ldots$. The fact that $\{X_n\}$ constitues a Markov chain and the fact that $\{S_n^j\}$, $j \in E$ constitutes renewal processes provide a justification for the terminology Markov renewal Process—as a generalisation of Markov chains and renewal processes. In view of proposition (3), an alternative definition of the Markov renewal process is also given below.

Definitions. If $N_j(t)$ denotes the number of transitions into state j (renewals of state j) in $[0, t]$, then

$$N(t) = \{N_1(t), N_2(t), \ldots, N_m(t)\}, \tag{2.12}$$

is called a *Markov renewal process.*

Suppose that the process starts with state i at time 0, i.e. $X_0 = Y(0) = i$.

Define

$$G_{ij}(t) = \Pr\{N_j(t) \geqslant 1 \mid Y(0) = i\} \tag{2.13}$$
$$= \Pr\{S_1 \leqslant t \mid Y(0) = i\},$$
$$f_{ij}(t) = \Pr\{Y(t) = j \mid Y(0) = i\} \tag{2.14}$$
$$Q_{ij}^{(n)}(t) = \Pr\{X_n = j, t_n \leqslant t \mid X_0 = i\}. \tag{2.15}$$

$G_{ij}(t)$ is the d.f. of the *first passage time* V_{ij} into state j given that the initial state is i. Here irrespective of the number of transitions (into other states) involved, the visit to j must be for the *first time*. G_{jj}, which gives the d.f. of the time between two occurrences of the state j, may be different from G_{ij} $(i \neq j)$. Denote $E(V_{ij}) = \mu_{ij}$, when it exists. A more detailed discussion on first passage time is given in a latter section.

Now $Q_{ij}^{(0)}(t) = 1$ or 0 depending on $j = i$ or not. Denote $Q^{(1)}$ by Q.

Proposition (4):
For $n = 0, 1, 2, \ldots$

$$Q_{ij}^{(n+1)}(t) = \sum_k \int_0^t \{dQ_{ik}(u)\}\, Q_{kj}^{(n)}(t-u). \tag{2.16}$$

This can be seen by considering that of the $(n+1)$ transitions, the first one takes place at some time u $(\leqslant t)$ and reaches some state k, and the subsequent n transitions from state k take place within time $t-u$, and end with j.

Markov renewal function. The functions

$$M(i, j, t) = E\{N_j(t) \mid X_0 = i\}, j \in E \tag{2.17}$$

are known as *Markov renewal functions* corresponding to Q.

Proposition (5): We have

$$M(i, j, t) = \sum_n Q_{ij}^{(n)}(t). \tag{2.18}$$

This can be seen by noting that

$$N_j(t) = \sum_n A_n, \text{ where}$$
$$A_n = 1, \text{ if } X_n = j, t_n \leqslant t$$
$$= 0, \text{ otherwise}$$

and $\Pr\{A_n = 1 \mid X_0 = i\} = \Pr\{X_n = j, t_n \leqslant t \mid X_0 = i\}$

$$= Q_{ij}^{(n)}(t). \tag{2.19}$$

For $j = i$, the interarrival times between two occurrences of i (having d.f. $G_{ii}(t)$) induce an ordinary renewal process. Denoting the n fold convolution of G_{ii} with itself by G_{ii}^{n*}, we can write the Markov renewal function (i.e. conditional expectation of the number of entrances to state i) as

$$M(i, i, t) = \sum_n G_{ii}^{n*}. \tag{2.20}$$

Proposition (6):

Blackwell's and Smith's renewal theorems hold for the (ordinary) Markov renewal function $M(i, i, t)$.

Classification of states, Suppose that the initial state in a Markov renewal process $\{X_n, t_n\}$ is i. The state i is said to be recurrent (transient) iff $G_{ii}(\infty) = 1 \; (< 1)$. A recurrent state is positive or null recurrent depending on $\mu_{ii} <$, or $= \infty$. We have the following result.

Proposition (7):

For the Markov renewal process $\{X_n, t_n\}$ state i is recurrent (transient) iff state i is recurrent (transient) in the embedded Markov chain $\{X_n\}$.

7.3 Markov Renewal Equation

We obtain a Markov renewal equation which gives a relationship between f_{ij} and Q_{ij}.

Theorem 7.1 For all i, j and for $t \geqslant 0$,

$$f_{ij}(t) = \delta_{ij}\, h_i(t) + \sum_k \int_0^t f_{kj}(t-x)\, dQ_{ik}(x), \tag{3.1}$$

where $$h_i(t) = 1 - \sum_k Q_{ik}(t) = 1 - W_i(t) = \Pr\{T_i > t\}.$$

Proof: We apply renewal argument; by conditioning on the time t_1 of the first transition as well as the state then entered, we get

$$f_{ij}(t) = \Pr\{Y(t) = j \mid X(0) = i\}$$

$$= \sum_k \int_0^\infty \Pr\{Y(t) = j \mid X_0 = i, X_1 = k, t_1 = x\}\, dQ_{ik}(x).$$

If $t_1 = x > t$, i.e. the first transition epoch is beyond t, then $Y(t) = Y(0) = i$. If $t_1 = x \leqslant t$, then starting from state k (reached at first transition by time x) the process ends up with state j by the remaining time $(t - x)$. Thus

$$\Pr\{Y(t) = j \mid X_0 = i, X_1 = k, t_1 = x\} = \delta_{ij}, \text{ if } x > t$$

Hence $$= f_{kj}(t-x), \text{ if } x \leqslant t.$$

$$f_{ij}(t) = \sum_k \left[\int_0^t f_{kj}(t-x)\, dQ_{ij}(x) + \delta_{ij} \int_t^\infty dQ_{ik}(x) \right].$$

Since
$$\sum_k \int_0^\infty dQ_{ik}(x) = 1,$$
the second term
$$= \delta_{ij}\{1 - \sum_k \int_0^t dQ_{ik}(x)\}$$
$$= \delta_{ij}\{1 - \sum_k Q_{ik}(t)\} = \delta_{ij}h_i(t) \tag{3.2}$$
and thus we get (3.1). ▲

The equation (3.1) in $f_{ij}(t)$ can be solved as a Markov renewal equation. We shall not consider this aspect here. We are interested mainly in the asymptotic behaviour of the stochastic process $Y(t)$, which will be found without explicitly finding the solution $f_{ij}(t)$ first. We shall introduce a matrix similar to the t.p.m. of a Markov chain.

Interval Transition Probability Matrix (i.t.p.m)

The matrix with functions of t, $k_{ij}(t)$, $i, j = 1, 2, \ldots, m$ as elements will be denoted by $\mathbf{K}(t)$, i.e. $\mathbf{K}(t) = (k_{ij}(t))$. The matrices obtained by taking derivatives, integrals, Laplace transforms etc. (provided they exist) of the elements will be represented accordingly. For example,
$$\int_0^t \mathbf{K}(x)dx = (\int_0^t k_{ij}(x)\,dx)$$
$$\mathbf{K}^*(s) = (k^*_{ij}(s)),$$
where $k^*(.)$ denotes Laplace transform.

We shall refer to these matrices here
$$\mathbf{F}(t) = (f_{ij}(t)),$$
$$\mathbf{C}(t) = (c_{ij}(t), \text{ where } c_{ij}(t)\,dt = p_{ij}dW_{ij}(t) = dQ_{ij}(t)$$
$$\mathbf{A}(t) = (a_{ij}(t)), \text{ where } a_{ij}(t) = \delta_{ij}\,h_i(t). \tag{3.3}$$
The matrix $\mathbf{A}$ is a diagonal matrix. Since $f_{ij}(t) \geqslant 0$ and $\sum_j f_{ij}(t) = 1$ for all $i = 1, 2, \ldots, m$, the matrix $\mathbf{F}(t)$ is a stochastic matrix; it is called *interval transition probability matrix* (i.t.p.m.).

Now (3.1) which holds for all states $i, j = 1, 2, \ldots, m$ can be put as
$$f_{ij}(t) = a_{ij}(t) + \int_0^t \sum_k c_{ik}(x) f_{kj}(t-x)\,dx, \quad i, j = 1, 2, \ldots, m \tag{3.4}$$
and in matrix form as
$$\mathbf{F}(t) = \mathbf{A}(t) + \int_0^t \mathbf{C}(x)\mathbf{F}(t-x)\,dx. \tag{3.5}$$
Denoting the convolution of $\mathbf{C}$ and $\mathbf{F}$ by $\mathbf{C}*\mathbf{F}$, we can put the same as

$$\mathbf{F}(\cdot) = \mathbf{A}(\cdot) + \mathbf{C}(\cdot) * \mathbf{F}(\cdot). \tag{3.6}$$

Taking Laplace transforms, we get

$$\mathbf{F}^*(s) = \mathbf{A}^*(s) + \mathbf{C}^*(s)\,\mathbf{F}^*(s). \tag{3.7}$$

Note that $\mathbf{C}$ is a somewhat special type of matrix.

7.4 Limiting Behaviour

Asymptotic behaviour of a stochastic process is a question of great general interest. We shall now investigate such behaviour in case of a semi-Markov process. Suppose that $\{X_n, t_n\}$ is a Markov renewal process and $Y(t) = X_n$, $t_n \leqslant t < t_{n+1}$, is the semi-Markov process induced by $Q_{ij}(t)$. The limiting distribution of the embedded Markov chain $\{X_n\}$ is not the same as that of the semi-Markov process. If the Markov chain is irreducible and ergodic and has the t.p.m. $\mathbf{P} = (p_{ij})$, then the limiting probabilities $\lim\limits_{n\to\infty} p_{ij}^{(n)} = v_j$, $j \in E$ are given as unique solutions of $v_j = \sum\limits_k v_k p_{kj}$, $j \in E$, i.e. the probability vector $\mathbf{V} = (v_1, v_2, \ldots)$ is the unique solution of $\mathbf{V} = \mathbf{VP}$.

In the ergodic theorem given below, the limiting distribution

$$\lim_{t\to\infty} f_{ij}(t) = \lim_{t\to\infty} \Pr\{Y(t) = j \mid Y(0) = i\}$$

is obtained. It also gives the relationship between the two limiting distributions.

Theorem 7.2 Suppose that $\{X_n\}$ is irreducible and ergodic and that the mean waiting time in state i, $M_i < \infty$ for any $i \in E$, then

$$\lim_{t\to\infty} f_{ij}(t) = f_{ij} = \frac{v_j M_j}{\sum\limits_k v_k M_k}. \tag{4.1}$$

Proof: Suppose that the variables T_{ij} are continuous having p.d.f. $w_{ij}(t)$. Then T_i is also continuous and has the p.d.f. $w_i(t) = \sum\limits_j p_{ij}\, w_{ij}(t)$. Referring to (3.3), we get

$$c_{ij}^*(s) = p_{ij}\, w_{ij}^*(s),\ a_{ij}^*(s) = \delta_{ij}\, h_i^*(s) = \delta_{ij}\left\{\frac{1}{s} - \frac{w_i^*(s)}{s}\right\}. \tag{4.2}$$

Taking limits as $s \to 0$,

$$c_{ij}^*(s) \to p_{ij},\ a_{ij}^*(s) \to \delta_{ij} \frac{d}{ds}(-w_i^*(s))\big|_{s=0} = \delta_{ij} M_i \tag{4.3}$$

so that as $s \to 0$

$$\mathbf{C}^*(s) = (c_{ij}^*(s)) \to (p_{ij}) = \mathbf{P} \tag{4.4}$$

and

$$\mathbf{A}^*(s) = (a_{ij}^*(s)) \to (\delta_{ij} M_i) = \mathbf{M};$$

$\mathbf{M}$ is a diagonal matrix with diagonal elements $M_1, M_2, \ldots$.

From (3 7) we get

$$[I - \mathbf{C}^*(s)]\,\mathbf{F}^*(s) = \mathbf{A}^*(s),$$

I being the unit matrix.

Thus
$$s\mathbf{F}^*(s) = \mathbf{B}(s)\,\mathbf{A}^*(s), \tag{4.5}$$
where
$$\mathbf{B}(s) = s\,[I - \mathbf{C}^*(s)]^{-1}. \tag{4.6}$$
Hence
$$\mathbf{B}(s)\,[I - \mathbf{C}^*(s)] = sI$$
or
$$\mathbf{B}(s) - \mathbf{B}(s)\mathbf{C}^*(s) = sI.$$
Denote
$$\lim_{s\to 0} \mathbf{B}(s) = \mathbf{B} = (b_{ij});$$
then taking limits as $s \to 0$, we get
$$\mathbf{B} - \mathbf{BP} = \mathbf{0}$$
or
$$\mathbf{B} = \mathbf{BP}. \tag{4.7}$$

Thus we find that $\mathbf{B}$ satisfies the same equation as $\mathbf{V}$, i.e.

$$b_{ij} = \sum_k b_{ik}p_{kj},\ j = 1, 2, \ldots \tag{4.8}$$

$$v_j = \sum_k v_k p_{kj},\ \ j = 1, 2, \ldots. \tag{4.9}$$

The solutions of (4.9) are unique. As both the sets (4.8) and (4.9) hold, the rows of $\mathbf{B}$ and $\mathbf{V}$ must be proportional, i.e.

$$b_{ij} = d_i v_j, \text{ where } d_i \text{ are constants.}$$

Now
$$f_{ij} = \lim_{t\to\infty} f_{ij}(t) = \lim_{s\to 0} sf_{ij}{}^*(s);$$
writing
$$\mathbf{F} = (f_{ij}) = \lim_{t\to\infty} (f_{ij}(t))$$
$$= \lim_{s\to 0} (sf_{ij}{}^*(s)) = \lim_{s\to 0} s\mathbf{F}^*(s),$$

we have from (4.5)
$$\mathbf{F} = \lim_{s\to 0} s\mathbf{F}^*(s) = \lim_{s\to 0} \mathbf{B}(s) \lim_{s\to 0} \mathbf{A}^*(s)$$
$$= \mathbf{BM}. \tag{4.10}$$

Thus
$$f_{ij} = b_{ij}M_j = d_i v_j M_j.$$
Since
$$\sum_j f_{ij} = 1,\ d_i = 1/\sum_j v_j M_j \text{ for all } i,$$

finally we get
$$\lim_{t\to\infty} f_{ij}(t) = f_{ij} = \frac{v_j M_j}{\sum_k v_k M_k}.$$

The limits $f_{ij}(t)$ are independent of the initial state i; we may write $f_{ij} = f_j$.

To complete the proof of the theorem, we note that in case the variables are discrete, by considering generating functions instead of Laplace transforms, and proceeding on the same lines, we can obtain the result. ▲

Particular case. When the waiting times are all exponential and $W_{ij}(t) = 1 - e^{-a_i t}$, then $Y(t)$ becomes a Markov process. We then get $M_j = 1/a_j$. Substituting this value of M_j in the expression for f_j, one can ultimately arrive at the expression (Eq. (5.19), Ch. 4).

$$\sum_j f_j a_{jk} = 0, \ (\sum_j f_j = 1)$$

satisfied by the limiting distribution $\{f_j\}$ of an irreducible Markov process

When the semi-Markov process reduces to an ordinary Markov chain, then $M_j = 1$ for all j and one gets $f_k = \nu_k$, as is to be expected. The relation $f_k = \nu_k$ holds also in the case when the unconditional waiting times in all the states (i.e. interoccurrence times between recurrence points) are independently distributed with the same mean $M_j = M$ for all j. It may be noted that the expression (4.1) giving f_{ij} involves only the ν_j's, the limiting transition probabilities of the embedded Markov chain and the mean waiting times $M_j = \sum_k p_{jk} \alpha_{jk}$, i.e. the mean α_{jk} and no other statistic of T_{jk}. The distinction between the quantities ν_j and f_j must be carefully observed. Suppose that a process has been in operation for a sufficiently long time and that a sufficiently large number of transitions have taken place so that a steady state has been reached. Suppose that the situation at some instant is considered. The quantity f_j gives the probability that the process is in (or occupies) the state j, while ν_j gives the probability that, given that the process is now making a transition, the same is to state j. Except in the cases stated above the two quantities ν_j, f_j are unequal. Even in the case when the process is a Markov process in continuous time they are unequal (see Example 4(b)). We consider a numerical example below.

Example 4(a). Suppose that the embedded Markov chain $\{X_n\}$ has the t.p.m.

$$P = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.4 & 0.4 \\ 0.1 & 0.5 & 0.4 \end{pmatrix}$$

and that the matrix of complementary distribution functions of T_{ij} is

$$W = \begin{pmatrix} e^{-2t} & e^{-5t} & e^{-t} \\ e^{-4t} & e^{-2t} & e^{-5t} \\ e^{-5t} & e^{-2t} & e^{-t} \end{pmatrix}.$$

Using $M_i = \sum_j p_{ij} \alpha_{ij}$, we get

$$M_1 = 0.51, \ M_2 = 0.33, \ M_3 = 0.67,$$

the limiting probabilities $\nu_j = \lim_{n \to \infty} p_{ij}^{(n)}$ have values

$$v_1 = 0.2353,\ v_2 = 0.4118,\ v_3 = 0.3529$$

(see Example 4(b) Ch. 3).

Using (4.1), we get

$$f_1 = 0.2437,\ \ f_2 = 0.2760,\ \ f_3 = 0.4803.$$

Example 4(b). Consider the $M/M/1$ queue discussed in Example 2(b). Suppose $\lambda < \mu$, and denote

$$a = \mu/(\lambda + \mu),\ b = \lambda/(\lambda + \mu);$$

then

$$p_{0,1} = 1,\ p_{j,j-1} = a,\ p_{j,j+1} = b, j \geqslant 1, p_{jk} = 0,\ k \neq j-1, j+1.$$

The elements of $\mathbf{V} = (v_0, v_1, v_2, \ldots), \Sigma\, v_i = 1$ are the solutions of $\mathbf{V} = \mathbf{V}P$. We have

$$v_0 = av_1$$

$$v_1 = v_0 + av_2$$

and
$$v_k = bv_{k-1} + av_{k+1},\ \ k = 2, 3, \ldots.$$

The solution of the last equation is given by

$$v_k = c_1 + c_2(b/a)^k,\ k = 1, 2, \ldots.$$

Solving in terms of v_0, we get

$$c_1 = 0 \text{ and } c_2 = v_0/b,$$

$$\begin{aligned} v_k &= (1/b)\,(b/a)^k\, v_0 \\ &= \left(\frac{\lambda+\mu}{\lambda}\right)\left(\frac{\lambda}{\mu}\right)^k v_0,\ \ k = 1, 2, \ldots \end{aligned}$$

Now
$$M_0 = 1/\lambda,\ M_k = 1/(\lambda + \mu),\ k = 1, 2, \ldots$$

so that

$$v_k M_k = \frac{v_0}{\lambda(1 - \lambda/\mu)}.$$

Thus we get

$$\begin{aligned} f_{ij} = f_j &= \frac{v_j M_j}{\sum\limits_k v_k M_k} \\ &= \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^j,\ \ j = 0, 1, 2, \ldots. \end{aligned}$$

This result obtained earlier, as solution of difference equations, is considered here as an illustration of application of the theory of s-M.P.

Note that
$$v_k > f_k,\ k \geqslant 1,\ v_0 < f_0.$$

Limiting Distribution of s-M.P. and 'Recurrence Times'

There are many other distributions connected with Markov renewal processes, whose asymptotic behaviour is of considerable significance and many limit theorems have been obtained. We shall discuss here the limiting behaviour in another important case.

Suppose that $\{X_n, t_n\}$ is a Markov renewal process and that $Y(t)$ is its associated s-M.P. For each t, there exists n, such that $t_n \leqslant t < t_{n+1}$; then $U(t) = t - t_n$, $V(t) = t_{n+1} - t$ are respectively called the time since the last transition and the time until the next transition ('recurrence times'). By applying renewal type of arguments and using the renewal theorems, the limiting form of the joint distribution of $Y(t)$, $U(t)$, $V(t)$ can be obtained as stated below. For details of proof, refer to Pyke (1961), Cinlar (1969) etc.

Theorem 7.3. If the process is irreducible and not lattice, then, as $t \to \infty$

$$\Pr\{Y(t)=j,\ X_{n+1}=k,\ U(t) \geqslant x,\ V(t) > y \mid Y(0)=i\}$$

$$\to \frac{v_j}{\sum\limits_r v_r M_r} \int\limits_{x+y}^{\infty} p_{jk}\{1 - W_{jk}(u)\}\, du, \tag{4.11}$$

provided the integral exists.

Summing over k, we have, as $t \to \infty$

$$\Pr\{Y(t)=j,\ U(t) \geqslant x,\ V(t) > y \mid Y(0)=i\}$$

$$\to \frac{v_j}{\sum\limits_r v_r M_r} \int\limits_{x+y}^{\infty} \{1 - \sum_k Q_{jk}(u)\}\, du = \frac{f_j}{M_j} \int\limits_{x+y}^{\infty} \Pr\{T_j > u\}\, du. \tag{4.12}$$

Thus, $\Pr\{U(t) \geqslant x,\ V(t) > y \mid Y(t)=j,\ Y(0)=i\}$

$$\to \int\limits_{x+y}^{\infty} \frac{\Pr\{T_j > u\}}{M_j}\, du. \tag{4.12a}$$

Putting $x=0$ in (4.11), we have, as $t \to \infty$,

$$\Pr\{Y(t)=j,\ X_{n+1}=k,\ V(t) > y \mid Y(0)=i\}$$

$$\to \frac{v_j}{\sum\limits_r v_r M_r} \int\limits_{y}^{\infty} p_{jk}\{1 - W_{jk}(u)\}\, du. \tag{4.13}$$

It may be noted that when the Markov renewal process reduces to an ordinary renewal process (i.e. there is only one state), the above (4.13) yields the limiting distribution of the residual lifetime, whereas (4.12a) gives the limiting joint distribution of residual and spent lifetimes.

7.5 First Passage Time

We consider here some of their properties. First, we obtain a relationship between $G_{ij}(t)$ and $f_{ij}(t)$.

Theorem 7.4. For all i, j and for $t \geqslant 0$,

$$f_{ij}(t) = \delta_{ij}\, h_i(t) + \int_0^t f_{jj}(t-x)\, dG_{ij}(x), \tag{5.1}$$

where $h_i(t) = \Pr\{T_i > t\}$.

Proof: The event $Y(t) = j$, given $Y(0) = i$ can happen in the following two mutually exclusive ways, which give the first and the second term respectively on the r.h.s. of (5.1):

(i) No transition occurs by time t; then $Y(t) = i$, and the time spent in state i exceeds t.

(ii) One or more transitions occur by time t. Suppose that the process reaches the state j for the first time by time $x < t$ and then starting from this initial state j, the process in the remaining time $t - x$, ends up again with state j. ▲

We now obtain a relation between $G_{ij}(t)$ and $Q_{ij}(t)$.

Theorem 7.5. For all states i, j and $t \geqslant 0$

$$G_{ij}(t) = Q_{ij}(t) + \sum_{k \neq j} \int_0^t G_{kj}(t-x)\, dQ_{ik}(x) \tag{5.2}$$

$$= Q_{ij}(t) + \sum_k \int_0^t G_{kj}(t-x)\, dQ_{ik}(x)$$

$$- \int_0^t G_{jj}(t-x)\, dQ_{ij}(x). \tag{5.2a}$$

Proof: The first term on the r.h.s of (5.2) corresponds to the case of direct transition from state i to state j in the first transition within t; the second term corresponds to the case of transition from state i to k $(\neq j)$ in the first transition at time $x < t$ and then starting from the initial state k, the process, in the remaining time $(t - x)$, ending up with state j. ▲

Laplace transform formulation. Suppose that the variables V_{ij}, T_{ij} are continuous and

$$dG_{ij}(x) = g_{ij}(x) dx \tag{5.3}$$

$$dQ_{ij}(x) = p_{ij}\, dW_{ij}(x) = p_{ij}\, w_{ij}(x)\, dx,$$

Then taking Laplace transforms of (5.1), we get

$$f_{ij}^*(s) = \delta_{ij}\, h_i^*(s) + g_{ij}^*(s) f_{jj}^*(s),$$

so that
$$f_{ii}^*(s) = \frac{h_i^*(s)}{1 - g_{ii}^*(s)},\ j = i \tag{5.4}$$

$$f_{ij}^*(s) = \frac{g_{ij}^*(s)\, h_j^*(s)}{1 - g_{jj}^*(s)},\ j \neq i. \tag{5.4a}$$

Taking Laplace transform of (5.2), we get

$$G^*_{ij}(s) = Q^*_{ij}(s) + \sum_{k\neq j} G^*_{kj}(s)\, q^*_{ik}(s). \tag{5.5}$$

Equivalently,

$$g^*_{ij}(s) = q^*_{ij}(s) + \sum_{k\neq j} g^*_{kj}(s)\, q^*_{ik}(s). \tag{5.6}$$

Denoting $(g^*_{ij}(s)) = G(s)$ and the diagonal matrix $(\delta_{ij} g^*_{ii}(s)) = G_j{}^*(s)$, (5.6) can be written in matrix form as

$$\mathbf{G}^*(s) = \mathbf{C}^*(s) + \mathbf{C}^*(s)\,\mathbf{G}^*(s) - \mathbf{C}^*(s)\,\mathbf{G}_j{}^*(s). \tag{5.6a}$$

Mean first passage time. If $G_{ij}(\infty) = 1$, then

$$\mu_{ij} = E(V_{ij}) = \int_0^\infty t\,dG_{ij}(t) \tag{5.7}$$

is the mean first passage time and μ_{jj} is the mean recurrrence time of state j.

Theorem 7.6. For all i, j

$$\mu_{ij} = M_i + \sum_{k\neq j} p_{ik}\mu_{kj}. \tag{5.8}$$

Proof: Suppose that the variables V_{ij}, T_{ij} are continuous and that the relations (5.3) hold; then $g^*_{ij}(0) = w^*_{ij}(0) = 1$

$$\mu_{ij} = -\frac{d}{ds} g^*_{ij}(s)\,|_{s=0}$$

$$\alpha_{ij} = -\frac{d}{ds} w^*_{ij}(s)\,|_{s=0}$$

and

$$M_i = \sum_k p_{ik}\,\alpha_{ik}.$$

From (5.6) we get

$$\frac{d}{ds} g^*_{ij}(s) = p_{ij}\frac{d}{ds} w^*_{ij}(s) + \sum_{k\neq j}\left\{p_{ik}\, w^*_{ik}(s)\frac{d}{ds} g^*_{kj}(s) + p_{ik}\, g^*_{kj}(s)\frac{d}{ds} w^*_{ik}(s)\right\}.$$

Hence

$$\begin{aligned}\mu_{ij} &= p_{ij}\,\alpha_{ij} + \sum_{k\neq j}\{p_{ik}\,\mu_{kj} + p_{ik}\,\alpha_{ik}\}\\ &= \sum_k p_{ik}\,\alpha_{ik} + \sum_{k\neq j} p_{ik}\,\mu_{kj}\\ &= M_i + \sum_{k\neq j} p_{ik}\,\mu_{kj}.\end{aligned}$$

The proof, when the variables are discrete can be given in the same manner by considering the generating functions of V_{ij}, T_{ij} in place of their Laplace transforms. ▲

Note: (1) The intuitive meaning of (5.8) is clear. The mean first passage time from state i to j is the meantime required to leave the state i *plus* the sum of the product of meantime required to go to state j from state k (for all k except $k = j$) reached at first transition and the probability that the first transition is from state i to k.

(2) Higher moments of V_{ij} can be obtained from (5.6). We obtain another interesting result.

Theorem 7.7. If the embedded chain $\{X_n\}$ is irreducible and ergodic, then for all i, j

$$\mu_{jj} = \sum_i v_i M_i / v_j. \tag{5.9}$$

Proof: Since $\{X_n\}$ is irreducible ergodic, v_i are given by unique solutions of

$$v_k = \sum_i v_i p_{ik}$$

Now multiplying both sides of (5.8) by v_i, $i = 1, 2, \ldots$ and adding, we get

$$\begin{aligned}\sum_i v_i \mu_{ij} &= \sum_i v_i M_i + \sum_i \{\sum_k v_i p_{ik} \mu_{kj} - v_i p_{ij} \mu_{jj}\} \\ &= \sum_i v_i M_i + \sum_k \{\sum_i v_i p_{ik}\} \mu_{kj} - \{\sum_i v_i p_{ij}\} \mu_{jj} \\ &= \sum_i v_i M_i + \sum_k v_k \mu_{kj} - v_j \mu_{jj}.\end{aligned}$$

Thus $$\mu_{jj} = \sum_i v_i M_i / v_j. \quad \blacktriangle$$

Note: We can write (5.9) as follows:

$$\mu_{jj} = \sum_i M_i (v_i / v_j).$$

The intuitive meaning of this expression would be clear once it is noted that (v_i/v_j) is the expected number of visits to i between two visits of j and each visit to i lasts a time whose mean is M_i.

Corollary 1. When the s-M.P. reduces to a simple Markov chain, then $M_j = 1$ for all j and then we get the well known result

$$\mu_{jj} = 1/v_j.$$

Corollary 2. From (4.1) and (5.9), we get

$$f_j = M_j / \mu_{jj}, \tag{5.10}$$

which gives f_j without directly involving v_i's.

It can be easily seen that the results derived for a finite set E will also hold good, in general, even for a set E containing a countably infinite number of points.

We have covered, within the scope of this book, the most important aspects of Markov renewal and semi-Markov processes. The niceties of theoretical structures and the vast potentialities of applications have attracted a large number of researchers to these topics. The literature has grown tremendously over the last two decades, from about 5 titles in 1960 to over 400 now. It is not possible to list them all here and reference may be made to Cinlar (1975) for a list of 70 titles and to the mimeographed compilation by Cheong *et al.* (1974) for an exhaustive list.

EXERCISES

7.1 The embedded Markov chain of a two-state {1, 2} s-M.P. has t.p.m.

$$P = \begin{pmatrix} 2/3 & 1/3 \\ 1/2 & 1/2 \end{pmatrix}$$

and the matrix of the complementary distribution functions of the waiting times is

$$W(t) = \begin{pmatrix} e^{-t} & e^{-3t} \\ e^{-2t} & e^{-4t} \end{pmatrix}.$$

Find (i) the L.T.'s of interval transition probabilities $f_{jk}(t)$, $j, k = 1, 2$

(ii) the limiting probabilities f_k, $k = 1, 2$.

7.2 For Problem 1, find

(i) the first passage time densities $g_{jk}(t)$,

(ii) the mean recurrence times of the states 1, 2, and

(iii) the mean of the first passage times from state j to state $k, j \neq k$.

7.3 Suppose that the employment status of certain section of labour population in an area can be described by a Markov chain with two states employed (0) and unemployed (1) and that the chain has t.p.m.

$$P = \begin{pmatrix} 0.4 & 0.6 \\ 0.7 & 0.3 \end{pmatrix}.$$

Suppose that the waiting times densities are given by $w_{00}(t) = e^{-t}$, $w_{01}(t) = 2e^{-2t}$, $w_{10}(t) = te^{-t}$ and $w_{11}(t) = \frac{1}{2}t^2 e^{-t}$.

(i) If a person has just been employed, what is the probability that he will be unemployed after time t?

(ii) If a person has just been employed, what is the p.d.f. of the first passage time to the state of unemployed? What is the mean of this first passage time?

(iii) What is the limiting probability that he will be: (a) employed and (b) unemployed after a sufficiently long time?

7.4 *Entrance probabilities.* Let $e_{ij}(t)\,dt$ denote the probability that the process will enter state j in the interval $(t, t + dt)$ given that it entered state i at $t = 0$. Show that

$$f_{ij}(t) = \int_0^t e_{ij}(x)\, h_j(t - x)\, dx,$$

where $$h_j(u) = \Pr\{T_j > u\}.$$

7.5 If the process is irreducible and ergodic, show that $\lim_{t\to\infty} e_{ij}(t) = e_j$

exists and is independent of the initial state i; the limit e_j (called the limiting entrance rate to state j) is given by

$$e_j = f_j/M_j$$

and further that $\quad e_j = 1/\mu_{jj}$.

7.6 Show that the second moments

$$E(V_{ij}^2) = \mu_{ij}^{(2)} = \int_0^\infty t^2 \, dG_{ij}(t)$$

of the first passage times V_{ij} (with d.f. G_{ij}) satisfy the relation

$$\mu_{ij}^{(2)} = M_i^{(2)} + \sum_{k \neq j} p_{ik} \{2\alpha_{ij}\mu_{kj} + \mu_{kj}^{(2)}\}.$$

Use this relation to find var (V_{12}) and var (V_{21}) for Exercise 7.1.

7.7 Show that for an irreducible and ergodic process, if ${}_jN_i$ denotes the number of visits to state i in between two visits to state j, then

$$E\{{}_jN_i\} = v_i/v_j$$

also $\quad E\{{}_jN_i\} = \mu_{jj}/\mu_{ii}$.

7.8 Show that for an irreducible ergodic process

$$\lim_{t\to\infty} \frac{M(k, j, t)}{M(h, i, t)} \to \frac{v_j}{v_i}.$$

7.9 For the $M/M/1$ queue (exponential interarrival and service times having mean $1/a$ and $1/b$ respectively) and with batch service (taking in a batch not more than c units) find the expressions for $Q_{ij}(t)$ and p_{ij}.

7.10 For the $M/G/1$ queue (with exponential interarrival time having mean $1/a$ and general service time with distribution $B(x)$ and mean $1/b$) taking t_n as the epoch of the nth departure and $Y(t)$ as the number left in the system by the most recent departure, find expressions for $Q_{ij}(t)$.

7.11 Repeat the above exercise with the modification that service is in batches of size not more than c (i.e. of size c or the number in the queue at the time of initiation of service, whichever is less).

7.12 Examine the validity of the relation $f_i = v_i$ in case of $GI/M/1$ and $M/G/1$ queues.

REFERENCES

D.J. Bartholomew, *Stochastic Models for Social Processes*, Wiley, New York (1967).

U. Narayan Bhat, *Elements of Applied Stochastic Processes*, Wiley, New York (1972).

C.K. Cheong, J.H.A. de Smit and J.L. Teugels, *Bibliography on semi-Markov processes* (mimeographed), 4th Ed., Inst. of Maths., Catholic Univ. of Louvain, Heverlee, (1974).

E. Cinlar, "Markov Renewal Theory", *Adv. Appl. Prob.*, **1**, 123–187 (1969).

———"Markov Renewal Theory: A survey", *Management Science*, **21**, 727–752 (1975)

A.J. Fabens, "The solution of queueing and inventory models by semi-Markov Processes", *J.R.S.S.B.* **23**, 113–127 (1961) (Correction, **24**, 455–456 (1962).

R.A. Howard, *Dynamic Probabilistic System*, Vol. II: *Semi-Markov and Decision Processes*, Wiley, New York (1971).

E.P.C. Kao, "Modelling the movement of coronary patients within a hospital by semi-Markov processes", *Opns. Res.*, **22**, 683–699 (1974).

N. Keyfitz, *Applied Mathematical Demography*, Wiley, New York (1977).

P. Lévy, "Processus semi-Markoviens", *Proc. Inter. Cong. Math.* (Amsterdam), **3**, 416–426 (1954).

M.F. Neuts, "The single server queue with Poisson input and semi-Markov service times", *J. Appl. Prob.*, **3**, 202–230 (1966).

R.G. Potter, "Births averted by contraception: An approach through renewal theory", *Theo. Pop. Biol*, **1**, 251–272 (1970).

R. Pyke, "Markov renewal processes: definitions and preliminary properties", *Ann. Math. Stat.*, **32**, 1231–1242 (1961).

———, "Markov renewal processes with finitely many states", *Ann. Math. Stat.*, **32**, 1243–1265 (1961).

S.M. Ross, *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco (1970).

M.C. Sheps, "On the time required for conception", *Population Studies* **18**, 85–97 (1964).

M.C. Sheps and J.A. Menken, *Mathematical Models of Conception and Birth*, Univ. of Chicago Press, Chicago (1973).

W.L. Smith, "Regenerative stochastic processes", *Proc. Roy. Soc.* A. **232**, 6–31 (1955).

CHAPTER 8

STATIONARY PROCESSES AND TIME SERIES

8.1 Introduction

A series of observations $x(t)$ made sequentially in time t constitutes a time series. Examples of data taken over a period of time are found in abundance in diverse fields such as meteorology, geophysics, biophysics, economics, commerce, communication engineering etc. Daily records of rainfall data, prices of a commodity etc. constitute time series. The variate t denotes time, i.e. changes occur in time. But this need not always be so. For example, the records of measurements of the diameter of a nylon fibre along its length (distance) t also give a time series. Here t denotes length.

The essential fact which distinguishes time series data from other statistical data is the specific order in which observations are taken. While observations from areas other than time series are statistically independent, the successive observations from a time series may be dependent, the dependence based on the position of the observation in the series. The time t may be a continuous or discrete variable. A general mathematical model of the time series $Y(t)$, $t \in T$ is as follows:

$$Y(t) = f(t) + X(t). \tag{1.1}$$

Here $f(t)$ represents the systematic part and $X(t)$ represents the random part. These two components are also known as *signal* and *noise* respectively. The model is theoretical; $f(t)$ and $X(t)$ are not separately observable. While the model for $Y(t)$ gives the structure of the generating process, a set of observations (or time series data) is a realisation or a sample function of the process. The effect of time may be in both the systematic and the random parts.

A general model in which the effect of time is represented in the random part involves a stochastic process, i.e. a random function of time. The systematic part is represented by a (non-random or deterministic) function of time. The important types of evolution (in time) of the systematic part are described by the following components.

(i) *Trend:* It is a slowly varying function of time due to growth and other such permanent effects. A trend is usually represented by means of a polynomial of a low degree or a short Fourier series.

(ii) *Cyclical component*: It is a broad long-run movement, more or less regular, caused by fluctuations in activity and is quasi-periodic in appearance.

(iii) *Seasonal component*: It is composed of regular variations, caused by changing seasons, rhythmic activities etc. Cyclic and seasonal movements are usually represented by suitable periodic functions. Trigonometric functions are generally used for representation of cyclic movements. Both the systematic and random parts may involve a number of unknown parameters. The two parts are considered to be additive.

For description and analysis of components involving the systematic part the reader may consult the excellent standard texts in time series, such as Kendall, Kendall and Stuart, Malinvaud, Anderson, Fuller etc.

We shall discuss here only the models, where the random part involves stationary stochastic processes (real-valued). Here we shall also confine mostly to processes in discrete time. A time series $\{x_t, t = 1, \ldots, m\}$ may be considered as a single particular realisation of the stationary stochastic process $\{X_t, t \in I\}$. We shall be concerned here with the stationary process $\{X_t, t \in I\}$ rather than with the analysis of the time series $\{x_t, t = 1, \ldots, m\}$.

As discussed in Ch. 2, a stochastic process $\{X_t, t \in I\}$ is said to be stationary in the wide-sense or covariance stationary iff, for all $t \in I$, the mean value function $E\{X_t\} = m$ is independent of t and the covariance function $E\{(X_t - m)(X_{t+k} - m)\} = C_k$ is a function of the difference (in lag) $((t+k) - t) = k$. The correlation function is $\rho_k = C_k/C_0$, where $C_0 = \text{var}(X_t)$. The graph of the correlation function is called the *correlogram* of the stochastic process.

Some examples of stationary processes were given in Ch. 2. We shall consider below some stationary processes which are used as models for the generation of the random part of a time series.

8.2 Models of Time Series

(a) Purely Random Process (or white noise process)

A completely random process $\{X_t\}$ has

$$E\{X_t\} = m, \tag{2.1}$$

which may be assumed to be 0, for simplicity, and

$$E\{X_t X_{t+k}\} = \begin{cases} \sigma^2, & k = 0 \\ 0, & k \neq 0. \end{cases} \tag{2.2}$$

The process is covariance stationary.

(b) First Order Markov Process

$\{X_t\}$ has the structure

$$X_t + \alpha_1 X_{t-1} = e_t, \quad |\alpha_1| < 1, \tag{2.3}$$

where $\{e_t\}$ is purely random process, say, with $m = 0, \sigma = 1$. Further

e_{t+1} is not involved in $X_t, X_{t-1}, \ldots$, i.e. X_t depends on $e_1, e_2, \ldots e_t$, but is independent of $e_{t+1}, e_{t+2}, \ldots$.

Multiplying both sides of (2.3) by X_{t-1} and taking expectation, we get

$$C_1 + \alpha_1 C_0 = 0$$

or, $$\rho_1 + \alpha_1 = 0, \quad \text{i.e.} \quad \rho_1 = -\alpha_1 = a \text{ (say)}, \ |a| < 1.$$

Proceeding in the same way, we get

$$\begin{array}{lll} \rho_2 + \alpha_1\rho_1 = 0, & \text{i.e.} & \rho_2 = (-\alpha_1)^2 = a^2 \\ \ldots & \ldots & \ldots \\ \rho_n + \alpha_1\rho_{n-1} = 0, & \text{i.e.} & \rho_n = (-\alpha_1)^n = a^n \\ \ldots & \ldots & \ldots \end{array}$$

The process is covariance stationary. The representation (2.3) is a linear stochastic difference equation; X_t can be expressed as a formal solution of the difference equation as follows.

Denote the backward shift operator by B, i.e. $BX_t = X_{t-1}$.

Then $$B^2X_t = X_{t-2} \quad \text{and} \quad B^kX_t = X_{t-k}, \ k = 1, 2, 3, \ldots.$$

Thus the difference equation can be written as

$$(1 + \alpha_1 B)\, X_t = e_t$$

or $$X_t = (1 + \alpha_1 B)^{-1} e_t = \sum_{k=0}^{\infty} \{(-\alpha_1)^k B^k\}\, e_t$$

$$= \sum_{k=0}^{\infty} \rho_1^k\, e_{t-k}. \tag{2.4}$$

(c) Moving Average (MA) Process

Here $\{X_t\}$ can be represented as

$$X_t = a_0 e_t + a_1 e_{t-1} + \ldots + a_h e_{t-h}, \tag{2.5}$$

where the a's are real constants and $\{e_t\}$ is a purely random process with mean 0 and variance σ^2. X_t can also be represented as

$$X_t = f(B)\, e_t, \quad \text{where} \quad f(B) = \sum_{r=0}^{h} a_r B^r;$$

when $a_h \neq 0$, $X_t, t \in T$ is called an MA process of order h.

We have $$E\{X_t\} = 0,$$

and $$\begin{aligned} C_k &= E\{X_t X_{t+k}\} \\ &= E\{a_0 a_k e_t^2 + a_1 a_{k+1} e_{t-1}^2 + \ldots + a_{h-k} a_h e_{t+k-h}^2 \\ &\qquad + \sum_{r \neq t} a_r a_t e_r e_t\} \\ &= (a_0 a_k + a_1 a_{k+1} + \ldots + a_{h-k} a_h)\, \sigma^2, \quad k \leqslant h \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\quad , \ k > h. \end{aligned}$$

A necessary but *not* sufficient condition that such a moving average representation of order h is possible is $C_k = 0$ for $k > h$. We have

$$\begin{aligned}\rho_k = \frac{C_k}{C_0} &= \frac{a_0 a_k + a_1 a_{k+1} + \ldots + a_{h-k} a_h}{a_0^2 + a_1^2 + \ldots + a_h^2}, \text{ if } h \geqslant k \\ &= 0, \text{ if } h < k.\end{aligned} \tag{2.6}$$

The process is covariance stationary.

If, in particular, $a_k = 1/(h+1)$, $k = 0, 1, \ldots, h$, then

$$\begin{aligned}\rho_k = \frac{h-k+1}{h+1} &= 1 - \frac{k}{h+1}, \ 0 \leqslant k \leqslant h \\ &= 0, \qquad \text{otherwise.}\end{aligned} \tag{2.7}$$

The correlogram consists of a linearly decreasing set of points.

The representation (2.5) of X_t consists of a linear expression with a finite number of terms in $e_t, e_{t-1}, \ldots, e_{t-h}$. It is said to be of order h. The first order Markov process (2.4) can be considered as a moving average of infinite order with $\rho_k = a^{|k|}$ for all k.

Wold (1954) has given a necessary and sufficient condition that there exists an MA model of order h corresponding to a specific set of autocorrelation coefficients. Since Wold's work, interest in this topic increased and several contributions have been made. Recently Davies *et al.* (1974) found general inequalities involving autocorrelations of an MA process.

(d) Autoregressive Process (AR Process)

The process $\{X_t\}$ given by

$$X_t + b_1 X_{t-1} + b_2 X_{t-2} + \ldots + b_h X_{t-h} = e_t, \ b_h \neq 0, \tag{2.8}$$

where $\{e_t\}$ is a purely random process, with mean 0, is called an *autoregressive process of order h*.

X_t can be obtained as a solution of the linear stochastic difference equation

$$g(B)\, X_t = e_t, \tag{2.9}$$

where

$$g(B) = \sum_{r=0}^{h} b_r B^r, \ b_0 = 1.$$

Suppose that $g(B) = \Pi\,(1 - z_i B)$, $z_i \neq z_j$, i.e. $z_1^{-1}, \ldots, z_h^{-1}$ are the distinct roots of the equation $g(z) = 0$. Further suppose that $|z_i| < 1$ for all i, i.e. all the roots of $g(z) = 0$ lie outside the unit circle, i.e. the roots z_i of the characteristic equation $f(z) \equiv \sum_{r=0}^{h} b_r z^{h-r} = 0$ (where $f(z) = z^{-h} g(z^{-1})$) all lie within the unit circle. The complete solution of (2.9) can be written as

$$X_t = \sum_{r=1}^{h} A_r z_r^t + \frac{1}{g(B)} e_t,$$

where A_r are constants. Now

$$\frac{1}{g(B)} e_t = \prod_{i=1}^{h} (1 - z_i B)^{-1} e_t$$

$$= \sum_{r=0}^{\infty} b_r' B^r e_t = \sum_{r=0}^{\infty} b_r' e_{t-r},\ b_0 = 1,$$

where b_r' are constants involving z_i's.

If the process is considered as begun long time ago, then the contribution of $\sum_{r=1}^{h} A_r z_r^t$ damps out of existence. X_t is then given by

$$X_t = \sum_{r=0}^{\infty} b_r' e_{t-r}. \tag{2.10}$$

Thus an AR process can be represented by an MA process of infinite order.

The coefficients b_r' of e_{t-r} in the right hand side of (2.10) can also be obtained as follows. Using the expressions for X_t as given in (2.10), for $t, t-1, \ldots, t-h$ and then substituting them in (2.8), we get

$$\Sigma b_r' e_{t-r} + b_1 \Sigma b_r' e_{t-1-r} + b_2 \Sigma b_r' e_{t-2-r} + \ldots + b_h \Sigma b_r' e_{t-h-r} = e_t.$$

Equating the coefficients of $e_t, e_{t-1}, \ldots$ from both sides, we get.

$$\left.\begin{aligned} b_0' &= 1 \\ b_1' + b_1 b_0' &= 0 \\ \ldots \quad & \quad \ldots \\ b_{h-1}' + b_1 b_{h-2}' + \ldots + b_{h-1} b_0' &= 0 \end{aligned}\right\} \tag{2.11}$$

and

$$b_r' + b_1 b_{r-1}' + \ldots + b_h b_{r-h}' = 0,\ r = h, h+1, \ldots. \tag{2.12}$$

In other words, b_r' satisfies the difference equation (2.12) together with the initial conditions (2.11). Thus, b_r' can be obtained by solving (2.1:) with the help of (2.11).

We may put the result in the form of a theorem as follows:

Theorem 8.1. If the roots of the equation

$$f(z) \equiv z^h + b_1 z^{h-1} + \ldots + b_h = 0$$

all lie within the unit circle, then the autoregressive process

$$X_t + b_1 X_{t-1} + \ldots + b_h X_{t-h} = 0,$$

can be represented as an infinite moving average

$$X_t = \sum_{r=0}^{\infty} b_r' e_{t-r},$$

where b_r' are the roots of the difference equation

$$b_r' + b_1 b_{r-1}' + \ldots + b_h b_{r-h}' = 0,\ r = h, h+1, \ldots$$

subject to the initial conditions

$$b_{,}'^{0} = 1,\ b_1' + b_1 b_1' = 0,\ b_2' + {}_1b\, b + b_2 b_0' = 0,\ \ldots,$$
$$b_{h-1}' + b_1 b_{a-2}' + \ldots + b_{h-1} b_0' = 0.$$

Let us now determine the autocorrelation function of X_t. We have $E\{X_t\} = 0$. Multiplying (2.8) by X_{t-k} and taking expectation, we find that the autocovariances satisfy the relation

$$C_k + b_1 C_{k-1} + \ldots + b_h C_{k-h} = 0,\ k = 1, 2, \ldots.$$

Noting that $\rho_k = C_k/C_0$,

we get

$$\rho_k + b_1\rho_{k-1} + \ldots + b_h\rho_{k-h} = 0,\ k = 1, 2, \ldots \tag{2.13}$$

or

$$g(B)\ \rho_k = 0,$$

which is a difference equation in ρ_k of order h. These are known as *Yule-Walker Equations.*

Now $\rho_0 = 1$, and putting $k = 1, 2, \ldots, h-1$ in (2.13), we get $(h-1)$ equations in $\rho_1, \ldots, \rho_{h-1}$ which can be solved in terms of the coefficients $b_1, \ldots, b_h$. These will constitute h initial conditions which could be used for the solution of the difference equation

$$\rho_k + b_1\rho_{k-1} + \ldots + b_h\rho_{k-h} = 0,\ k \geqslant h.$$

The difference equation in ρ_k $(k \geqslant h)$ can then be completely solved. Thus if the characteristic equation $f(z) = 0$, have distinct roots $z_1, \ldots, z_h$, then

$$\rho_k = \sum_{r=1}^{h} \alpha_r z_r{}^k, \tag{2.14}$$

where α_r are constants which can be determined from the initial conditions. The process is evidently covariance stationary.

Note: The process $\{X_t\}$ given by (2.8) cannot be represented by (2.10) when the roots of the characteristic equation $f(z) = 0$ do not all lie within the unit circle. In this case $E\{X_t\} = \Sigma\ A_r z_r{}^t$ which is a function of t does not damp out even for large t and the process is non-stationary.

Let us consider the case $h = 2$.

(e) Autoregressive Process of Order Two (Yule Process)

This is given by

$$X_t + b_1 X_{t-1} + b_2 X_{t-2} = e_t. \tag{2.15}$$

Assume that the roots z_1, z_2 of the characteristic equation

$$z^2 + b_1 z + b_2 = 0$$

are distinct, and that both the roots lie within the unit circle, i.e. $b_2 < 1$, and $4b_2 > b_1{}^2$. Neglecting the contribution $\sum_{r=1}^{\infty} A_r z_r{}^t$ for large t, we get

$$\begin{aligned} X_t &= [(1 - z_1 B)\,(1 - z_2 B)]^{-1}\, e_t \\ &= [1 + (z_1 + z_2)\, B + (z_1{}^2 + z_1 z_2 + z_2{}^2)\, B^2 + \ldots]\, e_t \\ &= \left[\frac{\sum_{r=0}^{\infty} (z_1^{r+1} - z_2^{r+1})}{(z_1 - z_2)}\, B^r\right] e_t \\ &= \sum_{r=0}^{\infty} b_r'\, e_{t-r}, \end{aligned} \tag{2.16}$$

where

$$\begin{aligned} b_r' &= (z_1^{r+1} - z_2^{r+1})/(z_1 - z_2), \text{ when } z_1 \neq z_2 \\ &= (r+1) z_1{}^r, \text{ when } z_1 = z_2. \end{aligned} \tag{2.17}$$

We get $E\{X_t\} = 0$ and

$$\operatorname{var}(X_t) = \Sigma (b_r')^2 \operatorname{var}(e_t)$$

$$= \frac{\operatorname{var}(e_t)}{(z_1 - z_2)^2} \cdot \left[\frac{1}{1 - z_1^2} - \frac{2}{1 - z_1 z_2} + \frac{1}{1 - z_2^2}\right]$$

so that
$$\frac{\operatorname{var}(X_t)}{\operatorname{var}(e_t)} = \frac{1 + b_2}{(1 - b_2)\{(1 + b_2)^2 - b_1^2\}}.$$

The Yule-Walker equations are $\rho_k + b_1 \rho_{k-1} + b_2 \rho_{k-2} = 0, k = 1, 2, \ldots,$

whence we get $\rho_k = \alpha_1 z_1^k + \alpha_2 z_2^k.$

For $k = 1$, $\rho_1 + b_1 + b_2 \rho_1 = 0,$

whence $\rho_1 = -\dfrac{b_1}{1 + b_2} = \dfrac{z_1 + z_2}{1 + z_1 z_2};$ further $\rho_0 = 1.$

Using these two values we can determine the two constants α_1 and α_2. Thus for unequal roots (real or complex) of the characteristic equation, i.e. for $z_1 \neq z_2$

$$\rho_k = \frac{z_1^{k+1}(1 - z_2^2) - z_2^{k+1}(1 - z_1^2)}{(z_1 - z_2)(1 + z_1 z_2)}, \quad k = 0, 1, 2, \ldots. \tag{2.18}$$

When the roots z_1, z_2 are complex, we can write $z_1 = pe^{i\theta}$, $z_2 = pe^{-i\theta}$, $0 \leqslant \theta \leqslant 2\pi$, so that $p = \sqrt{(z_1 z_2)} = \sqrt{b_2}$, $2p \cos\theta = z_1 + z_2 = -b_1$

or $\cos\theta = -b_1/2\sqrt{b_2}.$

Putting these expressions for z_1, z_2 in (2.18), we get

$$\rho_k = \frac{p^{k+1}\{e^{i\theta(k+1)}(1 - p^2 e^{-2i\theta}) - e^{-i\theta(k+1)}(1 - p^2 e^{2i\theta})\}}{p(1 + p^2)(e^{i\theta} - e^{-i\theta})}$$

$$= \frac{p^k}{1 + p^2}\left[\frac{2i \sin(k + 1)\theta - 2ip^2 \sin(k - 1)\theta}{2i \sin\theta}\right]$$

$$= p^k (\sin k\theta \cot\psi + \cos k\theta)$$

$$= p^k \frac{\sin(k\theta + \psi)}{\sin\psi}, \tag{2.19}$$

where $$\tan\psi = \frac{1 + p^2}{1 - p^2}\tan\theta. \tag{2.20}$$

For equal (and real) root of the characteristic equation, i.e. $z_1 = z_2 = z$,

$$\rho_k = \{1 + k(1 - z^2)/(1 + z^2)\} z^k, \quad k = 0, 1, 2, \ldots.$$

This process is clearly covariance stationary. As $p < 1$, $\rho_k \to 0$ for large k, The correlogram is oscillatory and decaying in character.

(f) Autoregressive Moving Average Process (ARMA process)

Consider a process having the representation

$$\sum_{r=0}^{p} b_r X_{t-r} = \sum_{s=0}^{q} a_s e_{t-s} \quad (b_0 = 1), \tag{2.21}$$

where $\{e_t\}$ is a purely random process with mean 0, and the roots of $f(z) = \sum_{r=0}^{p} b_r z^{p-r} = 0$ all lie within the unit circle. Such a process is called an *autoregressive moving average process* and is denoted by ARMA (p, q). We can write (2.21) as $g(B)X_t = h(B)\, e_t$, where $g(B)$ and $h(B)$ are polynomials in B of degrees p and q respectively. Then X_t can be represented as an infinite moving average $X_t = \sum_r v_r e_{t-r}$ as follows: Put $u_t = \sum_{s=0}^{q} a_s e_{t-s}$ and let the roots z_i, $i = 1, 2, \ldots, p$ of the equation $f(z) = 0$ be distinct. Then $X_t = \sum_{r=1}^{p} A_r z_r^t + \prod_{i=1}^{q} (1 - z_i B)^{-1} u_t$.

Now as $|z_i| < 1$, we have for large t (neglecting the terms $\sum A_r z_r^t$),

$$X_t = \prod_{i=1}^{p} (1 - z_i B)^{-1} u_t = \sum_{r=0}^{\infty} b_r' \, u_{t-r} = \sum_{r=0}^{\infty} b_r' \left(\sum_{s=0}^{q} a_s e_{t-r-s} \right)$$

$$= \sum_{r=0}^{\infty} v_r e_{t-r} \text{ (say).} \tag{2.22}$$

The v_r's, $r = 0, 1, 2, \ldots$ in (2.22) can be determined by substituting the above expression for X_t in the l.h.s. of (2.21) and then comparing the coefficients of e_{t-r} from both sides.

We have $E\{X_t\} = 0$. Multiplying (2.21) X_{t-k} $(k \geqslant 0)$ and taking expectations, we get

$$g(B)\rho_k = \sum_{r=0}^{p} b_r \rho_{k-r} = 0, \; k > q$$

whence
$$\rho_k = \sum_{r=1}^{p} \alpha_r z_r^k, \; k > q, \tag{2.23}$$

α_r being constants.

8.3 Time and Frequency Domain: Power Spectrum

Covariance and correlation functions of stationary processes are functions in the time domain. It is found that the study of processes in the 'time domain' is neither simple nor illuminating. An alternative approach for the study of stationary processes is through 'frequency domain'. This leads to simpler and more rewarding results. Consider a covariance stationary process in discrete time having C_k and ρ_k as covariance and correlation functions respectively.

Properties of Covariance and Correlation Functions

(1) The covariance (correlation) function is even, and positive semi-definite.

(2) The correlation function has the representation

$$\rho_k = \int_{-\pi}^{\pi} e^{ikw} \, dF_1(w), \tag{3.1}$$

where $F_1(w)$ has the character of

When F_1 is absolutely continuous (and $dF_1(w) = f_1(w)\,dw$), one gets the inverse representation

$$f_1(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \rho_k \, e^{-ikw}, \tag{3.2}$$

The corresponding representation of covariance function is

$$c_k = \int_{-\pi}^{\pi} e^{ikw} \, dF(w) \tag{3.3}$$

where $dF(w) = C_0 dF_1(w)$; when $dF(w) = f(w)\,dw$,

$$f(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} C_k \, e^{-ikw}. \tag{3.4}$$

The function $F_1(w)$ (so also $F(w)$) is called the *integrated spectrum*; the function $f(w)$ is known as the *spectral density function* and $f_1(w)$ as the *normalised spectral density function* of the process.

The information contained in the sequence $\{C_k\}$ is thus contained in the function $f(\lambda)$ and vice-versa. From covariances which are 'time domain' entities one gets a unique 'frequency domain' function.

The relation (3.1) is of the same form as that between characteristic function and distribution function of a r.v. As a consequence of non-negative definiteness of ρ_k, the relation (3.1) follows from a celebrated theorem due to Bochner-Khinchine.

Example 3(a). White noise process $\{X_t\}$. We have

$$E\{X_t\} = 0,$$
$$E\{X_r X_s\} = \sigma^2, \; r = s$$
$$= 0, \; r \neq s.$$

Thus $C_k = 0,\ k \neq 0,\ C_0 = \sigma^2$. We have

$$f(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} C_k \, e^{-ikw} = \frac{\sigma^2}{2\pi}, \; -\pi \leqslant w \leqslant \pi.$$

Again, given $f(w)$, we at once get C_k from the inverse relationship (3 3).

Autocovariance and Autocorrelation Generating Functions

The autocovariance generating function (*AGF*) of a stationary process $\{X_t\}$ is the power series

$$C(z) = \sum_{k=-\infty}^{\infty} C_k \, z^k,$$

(considered for values of z for which $C(z)$ is convergent).

Replacing C_k by ρ_k $(= C_k/C_0)$, one gets the autocorrelation generating function

$$S(z) = \sum_{k=-\infty}^{\infty} \rho_k \, z^k = C(z)/C_0 = C(z)/(\text{var } X_t).$$

The spectral density $f(w)$ can be expressed in terms of AGF as follows:

$$f(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} C_k\, e^{-ikw} = \frac{1}{2\pi} \{C(e^{-iw})\}$$

$$= \frac{\text{var } X_t}{2\pi} S\,(e^{-iw}).$$

Theorem 8.2. Let $Y_t = \sum_{j=0}^{\infty} a_j X_{t-j}$

where $\Sigma \mid a_j \mid < \infty$, be a stationary process. If $C_Y(z)$, $C_X(z)$ are the AGF's and $f_Y(w)$, $f_X(w)$ are the spectral density functions of $\{Y_t\}$ and $\{X_t\}$ respectively, then

$$C_Y(z) = h(z)\, h(z^{-1})\, C_X(z)$$

and $$f_Y(w) = h(e^{iw})\, h({}^{-iw})\, f_X(w),$$

where $$h(z) = \sum_{j=0}^{\infty} a_j z^j.$$

Proof: Assume that $E\,\{X_t\} = 0$, then $E\,\{Y_t\} = 0$. Denote $E\,\{X_t X_{t+k}\}$ by C_k and $E\,\{Y_t\, Y_{t+k}\}$ by R_k. We have

$$R_k = E\,\{Y_t Y_{t+k}\} = E\,\{(\sum_{j=0}^{\infty} a_j X_{t-j})\,(\sum_{l=0}^{\infty} a_l X_{t+k-l})\}$$

$$= \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} a_j a_l\, C_{k+j-l}$$

and hence

$$C_Y(z) = \sum_{k=-\infty}^{\infty} R_k z^k$$

$$= \sum_{k=-\infty}^{\infty} z^k \{ \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} a_j a_l\, C_{k+j-l}\}$$

$$= \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \sum_{k=-\infty}^{\infty} a_j a_l\, z^{-j+l}\, C_{k+j-l}\, z^{k+j-l}$$

$$= (\sum_{l=0}^{\infty} a_l z^l)\,(\sum_{j=0}^{\infty} a_j\, z^{-j})\,(\sum_{s=-\infty}^{\infty} C_s z^s),$$

(setting $s = k + j - l$)

$$= h(z)\, h(z^{-1})\, C_X(z).$$

Again

$$f_Y(w) = \frac{1}{2\pi} \{C_Y\,(e^{-iw})\}$$

$$= \frac{1}{2\pi} \{h(e^{-iw})\, h(e^{iw})\, C_X(e^{-iw})\}$$

$$= h(e^{iw})\, h(e^{-iw})\, f_X(w).$$ ▲

Note: Assume $\{X_t\}$ to be a white noise process; then the above theorem can be applied to obtain the spectral density functions of MA, AR and $ARMA$ processes.

Example 3(b). *MA* process of order $h: X_t = \sum_{j=0}^{h} a_j e_{t-j}$, $\{e_t\}$ being a white noise process having variance σ^2. We get

$$f_X(w) = \frac{\sigma^2}{2\pi}\{h(e^{iw})\, h(e^{-iw})\}$$

$$= (\sigma^2/2\pi)\{a_0 + a_1\, e^{iw} + \ldots + a_h\, e^{iwh}\}\{a_0 + a_1\, e^{-iw} + \ldots + a_h\, e^{-iwh}\}.$$

On simplification, we get

$$f_X(w) = (\sigma^2/2\pi) \sum_{k=-h}^{h} \{\sum_{r=0}^{h-k} a_r\, a_{r+k}\} \cos kw$$

$$= (\sigma^2/2\pi) \sum_{k=-h}^{h} C_k\, e^{-ikw},$$

where $C_k = E\{X_t\, X_{t+k}\}$, as is to be expected.

Example 3(c). First order autoregressive process or Markov process: $X_t = aX_{t-1} + e_t, |a| < 1$, $\{e_t\}$ being a white noise process having variance σ^2. The process can be represented as (see (2.4))

$$X_t = \sum_{j=0}^{\infty} a^j e_{t-j}.$$

We get

$$f_X(w) = \left(\frac{\sigma^2}{2\pi}\right)\{h\,(e^{iw})\, h\,(e^{-iw})\},$$

where

$$h\,(e^{iw}) = \sum_{j=0}^{\infty} a^j e^{iwj} = 1/(1 - ae^{iw}).$$

Thus $f_X(w) = \frac{1}{2\pi}\cdot\frac{\sigma^2}{(1 - ae^{iw})(1 - ae^{-iw})} = \frac{\sigma^2}{2\pi\,(1 - 2a\cos w + a^2)}$.

Example 3(d). AR process of order $h: X_t + b_1X_{t-1} + \ldots + b_hX_{t-h} = e_t$, where $\{e_t\}$ is a white noise process having variance σ^2 and the roots of the equation $f(z) = z^h + b_1z^{h-1} + \ldots + b_h = 0$ all lie within the unit circle. Applying the theorem to

$$e_t = \sum_{j=0}^{h} b_j X_{t-j},$$

we get

$$f(w) = (\sum_{j=0}^{h} b_j e^{ijw})(\sum_{k=0}^{h} b_k e^{-ikw}) f_X(w),$$

where $f(w)$ is the spectral density of $\{e_t\}$. But $f(w) = \sigma^2/2\pi$, so that we at once get

$$f_X(w) = \frac{\sigma^2}{2\pi}\,[(\sum_{j=0}^{h} b_j e^{ijw})(\sum_{k=0}^{h} b_k e^{-ikw})]^{-1}.$$

Particular case: $h = 2$. The spectral density function $f_X(w)$ of a second order autoregressive process

$$X_t + b_1X_{t-1} + b_2X_{t-2} = e_t$$

is given by

$$fx(w) = \frac{\sigma^2}{2\pi}[(1 + b_1e^{iw} + b_2e^{2iw})(1 + b_1e^{-iw} + b_2e^{-2iw})]^{-1}$$

$$= \left(\frac{\sigma^2}{2\pi}\right)\frac{1}{1 + b_1{}^2 + b_2{}^2 + 2b_1(1 + b_2)\cos w + 2b_2\cos 2w}.$$

Example 3(e). ARMA (p, q) process:

$$\sum_{r=0}^{p} b_rX_{t-r} = \sum_{r=0}^{q} a_re_{t-r}, \quad b_0 = 1,$$

where $\{e_t\}$ is a white noise process and the roots of the equation $f(z) = z^h + b_1z^{h-1} + \ldots + b_h = 0$ all lie within the unit circle. Putting

$$Z_t = \sum_{r=0}^{q} a_re_{t-r}$$

and applying the theorem, we get

$$f_z(w) = \frac{\sigma^2}{2\pi}\left(\sum_{r=0}^{q} a_re^{irw}\right)\left(\sum_{s=0}^{q} a_se^{-isw}\right),$$

where $fz(w)$ is the spectral density of the process $\{Z_t\}$. Again from $Z_t = \sum_{r=0}^{p} b_rX_{t-r}$, we get

$$fz(w) = \left(\sum_{r=0}^{p} b_re^{irw}\right)\left(\sum_{s=0}^{p} b_se^{-isw}\right) fx(w).$$

Thus equating the two expressions for $fz(w)$, we get

$$fx(w) = \frac{\sigma^2}{2\pi}\cdot\frac{\left(\sum_{r=0}^{q} a_re^{irw}\right)\cdot\left(\sum_{s=0}^{q} a_se^{-isw}\right)}{\left(\sum_{r=0}^{p} b_re^{irw}\right)\cdot\left(\sum_{s=0}^{p} b_se^{-isw}\right)}.$$

Continuous Parameter Processes

Stationarity for a process in continuous time is defined in a similar manner. A process $\{X(t), -\infty < t < \infty\}$ in continuous time is wide sense stationary iff, for all t, the mean value function $E\{X(t)\} = m$ is independent of t and the covariance function $E[\{X(t) - m)\}\{(X(t + v) - m\}] = R(v)$ is a function of the lag difference $(t + v) - t = v$. The continuous time analogue of (3.3) is given below.

$R(v)$ is positive semidefinite and has the representation

$$R(v) = \int_{-\infty}^{\infty} e^{ivw}\, dF(w), \tag{3.5}$$

where F is nondecreasing and bounded; when $dF(w) = f(w)\, dw$,

$$f(w) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ivw} R(v)\, dv. \tag{3.6}$$

[The result follows from Bochner's theorem; for its proof see Goldberg (1961)].

Example 3(f). For a process having $R(v) = ae^{-b|v|}$, the spectral density function is

$$
\begin{aligned}
f(w) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iwv}\, ae^{-b|v|}\, dv \\
&= \frac{a}{2\pi} \left[\int_{-\infty}^{0} e^{(b-iw)v}\, dv + \int_{0}^{\infty} e^{-(b+iw)v}\, dv \right] \\
&= \frac{a}{2\pi} \left(\frac{1}{b-iw} + \frac{1}{b+iw} \right) \\
&= \frac{ab}{\pi (b^2 + w^2)}.
\end{aligned}
$$

8.4 Statistical Analysis of Time Series: Some Observations

Time series is one of the most widely pursued areas of statistics. Problems of statistical inference in time series, though very important from the point of view of practitioners, are beyond the scope of this book. However, we propose to give here very briefly an idea of the type of problems that arise.

The special feature is that the data are arranged with respect to time and successive observations are considered to be not independent. The problem of inference of such types of stochastic processes arising in diverse fields does not differ in principle from that of other types of statistical data; but the existence of some dependence in successive observations often implies that the classical methods become inadequate and need adequate extension. For example, the asymptotic theory of maximum likelihood estimates does not apply in general to dependent observations. Further, independent repetitions of the process are not available; instead a single observed sequence is available and this also makes new methods necesssary. Difficulty arises because of the fact that dependence has so many more possibilities *a priori* than independence, and more often than not, the distribution and the dependence of the observations representing stochastic processes may not be clearly known. The sampling theory of time series, like that of stochastic processes in general, is still in a comparatively early stage of development, and most of the results concern large samples, and small sample methods are less often available.

The inference problems centering the parametric models are concerned with the parameters of the models. For example, for the AR model

$$\sum_{r=0}^{h} b_r X_{t-r} = e_t,\ b_0 = 1,$$

the problems include estimation of the b's, the variance of e_t, the distribution of the estimates and hypotheses testing and similarly for the MA, ARMA models. There has been considerable amount of work on estimation and the question is still engaging the attention of researchers (see Chatfield (1977), for a review of recent developments and for an exhaustive list of references).

Nonstationary models have also been extensively studied (e,g. Priestley, Granger *et al*, Koopmans *et al*; (also see Box and Jenkins (1976) for building of stochastic models, specially non-stationary and their use in important areas of application).

In addition to making inferences about the statistical structure of the underlying process, the analysis of time series has two other purposes: forecasting and control. The Box-Jenkins forecasting method, based on a class of models called integrated autoregressive moving average (ARIMA) processes, has roused keen interest. They start with a modified type of ARMA model, such as $g(B)X_t = h(B)\, e_t$, where some roots (say, $d \geqslant 1$) of $g(B)$ are equal to unity and the others lie outside the unit circle. The corresponding nonstationary process (called ARIMA process) has the representation

$$f(B)\,(1-B)^d\, X_t = h(B)\, e_t,$$

where all the roots of $f(B)$ lie outside the unit circle. Introducing seasonality components, Box and Jenkins (pp. 303–305) arrive at the general multiplicative seasonal model

$$f_p(B)\, F_P(B^s)\,(1-B)^d\,(1-B^s)^D X_t = h_q(B)\, H_Q(B^s)\, e_t,$$

where $\{e_t\}$ is a white noise process, f_p, h_q are polynomials in degree p, q in the lag operator B, F_P, H_Q are polynomials in B^s, s being the number of seasons per year. It is assumed that f_p, h_q have all zeros outside the unit circle, h_q has no zeros on the unit circle and all the polynomials f_p, h_q, F_P, H_Q have unity as the constant terms. This class provides a wide range of models, both *stationary* and *non-stationary*. Box and Jenkins claim that this class is sufficiently flexible to provide a close approximation to the true model for many sets of homogeneous data. The main problem is the selection of an appropriate model. Box-Jenkins provide detailed techniques for identification of a proper model, for estimation of its parameters and for checking the adequacy of the model. Their techniques have aroused keen interest resulting in further active work.

Apart from the estimation and associated problems of a parametric model there are some other statistical problems associated with a stationary time series with mean μ. Given a consecutive sample of observations $X_1, \ldots, X_T$, one is concerned with the estimation of the mean, of autocovariances or autocorrelations, of the spectral density and also with the large sample behaviour of the estimates (see Anderson (1971), Fuller (1976) for a detailed account). The problem of estimation of the spectral density function has long attracted the attention of time series researchers (Bartlett, Blackman and Tukey, Parzen, Grenander, Medhi, Priestley, Rao etc., to name a few; see Jenkins and Watts (1968), Hannan (1970) for some details). The difficulties of selecting a best estimator on theoretical grounds are still being appreciated (Bloomfield, 1976).

EXERCISES

8.1 Do the autocorrelation functions uniquely determine a moving average process? Discuss with the help of some suitable examples.

8.2 Verify that the four moving average processes

$$X_t = 1.6\, e_t - 0.8\, e_{t-1} + 0.4\, e_{t-2} - 0.2\, e_{t-3}$$

$$X_t = -0.2\, e_t + 0.4\, e_{t-1} - 0.8\, e_{t-2} + 1.6\, e_{t-3}$$

$$X_t = 0.4\, e_t - 0.2\, e_{t-1} + 1.6\, e_{t-2} - 0.8\, e_{t-3}$$

$$X_t = -0.8\, e_t + 1.6\, e_{t-1} - 0.2\, e_{t-2} + 0.4\, e_{t-3}$$

all have the first three autocorrelation coefficients $-42/85$, $4/17$ $-8/85$ and the rest zero (Wold, 1954).

8.3 Find two moving average processes which have for autocorrelation $\rho_0 = 1$, $\rho_1 = .4$, $\rho_k = 0$, $k \geqslant 2$.

8.4 A set of autocorrelation functions $\rho_0 = 1, \rho_1, \ldots, \rho_h$ of an autoregressive process are given. Will these uniquely determine the autoregressive process?

It is given that an autoregressive process of order two has autocorrelation coefficients $\rho_0 = 0$, $\rho_1 = a$, $\rho_2 = b$. Determine the autoregressive process.

8.5 Consider the ARMA (1, 1) model

$$X_t - a_1 X_{t-1} = e_t + b_1 e_{t-1},\ t = 0, \pm 1, \pm 2, \ldots$$

where $|a_1| < 1$, $E(e_t) = E(e_r^2) = \sigma^2$, and $E(e_r e_s) = 0$, $r \neq s$.

Show that the model can be represented as an infinite moving average

$$X_t = e_t + (b_1 + a_1) \sum_{k=1}^{\infty} a_1^{k-1} e_{t-k}.$$

8.6 Let X_t be an MA process of order h given by

$$X_t = a_0 e_t + a_1 e_{t-1} + \ldots + a_h e_{t-h},\ a_h \neq 0,$$

where $\{e_t\}$ is a purely random process. If the roots of the characteristic equation

$$z^h + a_1 z^{h-1} + \ldots + a_h = 0$$

all lie within the unit circle, then X_t can be expressed as an autoregressive process of infinite order

$$\sum_{r=0}^{\infty} c_r X_{t-r} = e_t,$$

where the coefficients c_r satisfy the difference equation

$$c_r + a_1 c_{r-1} + \dots + a_h c_{r-h} = 0, r = h, h+1 \dots,$$

and the initial conditions

$$c_0 = 1, c_1 + a_1 = 0, c_2 + a_1 c_1 + a_2 = 0, \dots,$$

$$\dots, c_{h-1} + a_1 c_{h-2} + \dots + a_{h-1} = 0$$ (Fuller, 1976).

8.7 Let X_t be an ARMA (p, q) process given by

$$g(B)X_t = f(B) e_t$$

where $\{e_t\}$ is a purely random process,

$$g(B) = \sum_{r=0}^{p} b_r B^r, b_0 = 1,$$

$$f(B) = \sum_{r=0}^{q} a_r B^r, a_0 = 1$$

and the roots of $g(1/z) = 0$ all lie outside the unit circle. Show that X_t can then be represented as an *MA* process of infinite order

$$X_t = \sum_{r=0}^{\infty} c_r e_{t-r}.$$

8.8 Consider the process

$$X_t = aX_{t-1} + b + e_t,$$

where $\{e_t\}$ is a purely random process with variance unity. Further assume that the process was started at time 0 and that $X_0 = 0$. Show that

$$X_t = \left(\frac{1-a^t}{1-a}\right) b + \sum_{s=1}^{t} a^{t-s} e_s$$

Also show that

$$E\{X_t\} = \left(\frac{1-a^t}{1-a}\right) b$$

and $$C_k = a^k \left(\frac{1-a^{2t}}{1-a^2}\right).$$

Deduce that, if the series was started long ago, and if $|a| < 1$

$$C_k = a^k/(1-a^2).$$

8.9 Let $$X_t + a_1 X_{t-1} + a_2 X_{t-2} = Y_t,$$

where $$Y_t + b_1 Y_{t-1} + b_2 Y_{t-2} = e_t,$$

and $\{e_t\}$ is a purely random process and let the roots of $m^2 + a_1 m + a_2 = 0$ and $n^2 + b_1 n + b_2 = 0$ be less than unity in absolute value. Obtain the spectral density function of $\{X_t\}$.

8.10 Let $Z_t = X_t + Y_t$, where $\{X_t\}$ and $\{Y_t\}$ are two stationary processes with spectral density $f_X(w)$ and $f_Y(w)$ respectively. Then show that $\{Z_t\}$ is stationary and has spectral density $f_Z(w) = f_X(w) + f_Y(w)$. Hence find the spectral density of

$$Z_t = X_t + e_t,$$

where $X_t + aX_{t-1} = v_t$, and $\{e_t\}$ and $\{v_t\}$ are two independent white noise processes.

REFERENCES

T.W. Anderson, *The Statistical Analysis of Time Series,* Wiley, New York, (1971).

M.S. Bartlett, *An Introduction to Stochastic Processes*, Camb. Univ. Press London, (1954).

M.S. Bartlett and J. Medhi, "On the efficiency of procedures for smoothing periodograms with continuous spectra", *Biometrika*, **42**, 143–150 (1955).

R.B. Blackman and J.W. Tukey, *The Measurement of Power Spectra,* Dover, New York, (1959).

P. Bloomfield, *Fourier Analysis of Time Series: An Introduction*, Wiley, New York, (1976).

G.E.P. Box and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco (1970); Revised ed. (1976).

C. Chatfield, "Some recent developments in time series analysis", *J.R.S.S.A.*, **140**, 492–510 (1977).

N. Davies, M.B. Pata and M.G. Frost, "Maximum autocorrelations for moving average processes", *Biometrika*, **66**, 295–336 (1974).

W.A. Fuller, *Introduction to Statistical Time Series,*Wiley, New York, (1976).

R.R. Goldberg, *Fourier Transform*, Camb. Univ. Press, London (1961).

C.W.J. Granger and M. Hatanka, *Spectral Analysis of Economic Time Series*, Princeton Univ. Press, Princeton, N.J. (1964).

C.W.J. Granger, M. Hatanka and M.J. Morris, "Time Series modelling and interpretation", *J.R.S.S.A.* **139**, 246–257 (1976).

U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, Wiley, New York, (1957).

E.J. Hannan, *Time Series Analysis*, Methuen, London, (1960).

———, *Multiple Time Series,* Wiley, New York, (1970).

M.G. Kendall. *Time Series*, C. Griffin, London, (1974).

M.G. Kendall and A. Stuart, *The Advanced Theory of Statistics,* Vol. 3, C. Griffin, London, (1966).

G.M. Jenkins and D.G. Watts, *Spectral Analysis and its Application*, Holden-Day, San Francisco, (1968).

L.H. Koopmans, *The Spectral Analysis of Time Series*, Academic Press, New York, (1974).

B. Malinvaud, *Statistical Methods in Econometrics*, North Holland, Amsterdam, (1970).

J. Medhi, "A note on the properties of a test procedure for discrimination between two types of spectra of a stationary process," *Skand. Akt.* **20**, 6–13 (1959).

J. Medhi and T. Subba Rao, "Sequential and non-sequential test procedures for discrimination between discrete, continuous and mixed spectra of a stationary time series", *J. Ind. Stat. Ass.*, **5**, 1-10 (1967).

E. Parzen, "An approach to time series Analysis", *Ann. Math. Stat.*, **32**, 951-989 (1961).

M B. Priestley, "The analysis of stationary processes with mixed spectra, I & II" *J.R.S.S.B.*, **24**, I: 215–233, II: 511–523 (1962).

———, "Evolutionary spectra and nonstationary process", *J.R.S.S.B.*, **27**, 204–237 (1965).

H. Tong and M.B. Priestley, "On the analysis of bivariate nonstationary processes," *J.R.S.S.* Series **B**. 153–166 (1973).

T. Subba Rao, "Problems of statistical analysis in Time Series," Doctoral thesis, Gauhati University, India, (1966).

———, "On the cross periodogram of a stationary Gaussian vector process," *Ann. Math. Stat.* **38**, 593–597 (1967).

H. Wold, *A Study in the Analysis of Stationary Time Series*, 2nd ed. Almqvist & Wiksells, Stockholm, (1954).

———, *Bibliography on Time Series and Stochastic Process*, Oliver and Boyd, Edinburgh, (1965).

CHAPTER 9

BRANCHING PROCESSES

9.1 Introduction

The history of the study of branching processes dates back to 1874 when a mathematical model was formulated by Galton and Watson for the problem of 'extinction of families'. The model did not attract much attention for a long time; the situation gradually changed and during the last 40 years much attention has been devoted to it. This is because of the development of interest in the applications of probability theory, in general, and also because of the possibility of using the models in a variety of biological, physical, and other problems where one is concerned with objects (or individuals) that can generate objects of similar kind; such objects may be biological entities, such as human beings, animals, genes, bacteria and so on, which reproduce similar objects by biological methods or may be physical particles such as neutrons which yield new neutrons under a nuclear chain reaction or in the process of nuclear fission.

We consider first the discrete time case. Suppose that we start with an initial set of objects (or individuals) which form the 0-th generation—these objects are called *ancestors*. The offsprings reproduced or the objects generated by the objects of the 0-th generation are the 'direct descendants' of the ancestors, and are said to form the 1st generation; the objects generated by these of the 1st generation (or the direct descendants of the 1st generation) form the 2nd generation, and so on, the direct descendants of the rth generation form the $(r+1)$st generation. The number of objects of the rth generation $(r = 0, 1, 2, \ldots)$ is a random variable. We assume that the objects reproduce independently of other objects, i.e. there is no interference.

Definition. Let the random variables $X_0, X_1, X_2, \ldots$ denote the size of (or the number of objects in) the 0th, 1st, 2nd, ..., generations respectively. Let the probability that an object (irrespective of the generation to which it belongs) generates k similar objects be denoted by p_k, where $p_k \geqslant 0$, $k = 0, 1, 2, \ldots, \sum_k p_k = 1$.

The sequence $\{X_n, n = 0, 1, 2, \ldots\}$ constitutes a *Galton-Watson branching process* (or simply a *G.W. branching process*) with *offspring distribution* $\{p_k\}$.

Our interest lies mainly in the probability distribution of X_n and the probability that $X_n \to 0$ for some n, i.e. the probability of ultimate extinction of the family.

Note 1: Unless otherwise stated, we shall assume that $X_0 = 1$, i.e. the process starts with a single ancestor.

Note 2: The sequence $\{X_n\}$ forms a Markov chain with transition probabilities

$$p_{ij} = \Pr\{X_{n+1} = j \mid X_n = i\},\ i, j = 0, 1, 2, \ldots.$$

It is however not always easy to specify p_{ij}.

Note 3: The generating functions prove very useful in the study of branching processes.

9.2 Properties of Generating Functions of Branching Processes

We have

$$X_{n+1} = \sum_{r=1}^{X_n} \zeta_r, \tag{2.1}$$

where ζ_r are i.i.d. random variables with distribution $\{p_k\}$. Let

$$P(s) = \sum_k \Pr\{\zeta_r = k\} s^k = \sum_k p_k s^k \tag{2.2}$$

be the p.g.f. of $\{\zeta_r\}$ and let

$$P_n(s) = \sum_k \Pr\{X_n = k\} s^k,\ n = 0, 1, 2, \ldots \tag{2.3}$$

be the p.g.f. of $\{X_n\}$.

We assume that $X_0 = 1$; clearly $P_0(s) = s$ and $P_1(s) = P(s)$. The r.v.'s X_1 and ζ_r both give offspring distribution.

Theorem 9.1. We have

$$P_n(s) = P_{n-1}(P(s)) \tag{2.4}$$

and

$$P_n(s) = P(P_{n-1}(s)). \tag{2.5}$$

Proof: We have

$$\Pr\{X_n = k\} = \sum_{j=0}^{\infty} \Pr\{X_n = k \mid X_{n-1} = j\} \cdot \Pr\{X_{n-1} = j\}$$

$$= \sum_{j=0}^{\infty} \Pr\{\sum_{r=1}^{j} \zeta_r = k\} \cdot \Pr\{X_{n-1} = j\}$$

so that, for $n = 1, 2, \ldots$

$$P_n(s) = \sum_{k=0}^{\infty} \Pr\{X_n = k\} s^k$$

$$= \sum_{k=0}^{\infty} s^k \left[\sum_{j=0}^{\infty} \Pr\{\sum_{r=1}^{j} \zeta_r = k\} \Pr\{X_{n-1} = j\}\right]$$

$$= \sum_{j=0}^{\infty} \Pr\{X_{n-1} = j\} \left[\sum_{k=0}^{\infty} \Pr\{\zeta_1 + \zeta_2 + \ldots + \zeta_j = k\} s^k\right].$$

The expression within square brackets being the p.g.f. of the sum $\zeta_1 + \ldots + \zeta_j$ of j i.i.d. random variables each with p.g.f. $P(s)$, equals $[P(s)]^j$. Thus

$$P_n(s) = \sum_{j=0}^{\infty} \Pr\{X_{n-1} = j\} [P(s)]^j$$

$$= P_{n-1}(P(s)).$$

Thus we get (2.4). Putting $n = 2, 3, 4, \ldots$, we get

$$P_2(s) = P_1(P(s)) = P(P(s)),\ P_3(s) = P_2(P(s)),\ P_4(s) = P_3(P(s))$$

and so on. Iterating (2.4) we get

$$\begin{aligned} P_n(s) &= P_{n-1}(P(s)) = P_{n-2}(P(P(s))) \\ &= P_{n-2}(P_2(s)). \end{aligned} \tag{2.6}$$

For $n = 3$, $P_3(s) = P_1(P_2(s)) = P(P_2(s))$.

Again iterating (2.6), we get

$$P_n(s) = P_{n-3}(P(P_2(s)) = P_{n-3}(P_3(s)),$$

and for $n = 4$, $\quad P_4(s) = P_1(P_3(s)) = P(P_3(s))$.

Thus $\quad P_n(s) = P_{n-k}(P_k(s)),\ k = 0, 1, 2, \ldots, n,$

and for $k = n - 1$,

$$P_n(s) = P_1(P_{n-1}(s)) = P(P_{n-1}(s)).$$

Thus we get (2.5). ▲

Note that even when $X_0 = i \neq 1$, the relation (2.5) holds but (2.4) does not hold.

Moments of X_n

Theorem 9.1 could profitably be used to find the moments of X_n. We have $P'(1) = E(\zeta_r) = E(X_1) = m$ (say).

Theorem 9.2. If $m = E(X_1) = \sum_{k=0}^{\infty} kp_k$, and $\sigma^2 = \text{var}(X_1)$ then

$$E\{X_n\} = m^n \tag{2.7}$$

and

$$\begin{aligned} \text{var}(X_n) &= \frac{m^{n-1}(m^n - 1)}{m - 1}\sigma^2, \text{ if } m \neq 1 \\ &= n\sigma^2 \qquad\qquad\quad , \text{ if } m = 1. \end{aligned} \tag{2.8}$$

Proof: Differentiating (2.4), we get

$$P_n'(s) = P_{n-1}'(P(s))P'(s)$$

whence $\quad P_n'(1) = P_{n-1}'(1)P'(1) = mP_{n-1}'(1)$

and on iterating

$$\begin{aligned} P_n'(1) &= m^2 P_{n-2}'(1) \\ &\vdots \\ &= m^{n-1}P'(1) = m^n. \end{aligned}$$

Thus $\quad E(X_n) = P_n'(1) = m^n.$

Differentiating (2.5) twice and proceeding in a similar fashion, one can find the second moment $P_n''(1)$, and thus the variance of X_n in the form (2.8). ▲

One can likewise proceed to get higher moments of X_n.

Alternatively, the mean and the variance of X_n can be obtained by noting that X_{n+1} is the sum of a random number of i.i.d. random variables, and using standard formulas.

Let $m \neq 1$. We can use the Corollary to Theorem 1.3 to find $E(X_{n+1})$. Since $X_{n+1} = \sum_{r=1}^{X_n} \zeta_r$, we have

$$E(X_{n+1}) = E(\zeta_r)\, E(X_n) = mE(X_n).$$

The solution of the difference equation is given by

$$E(X_n) = Cm^n,\ n = 1, 2, 3, \ldots.$$

Since $E(X_1) = E(\zeta_r) = m,\ C = 1$. Thus $E(X_n) = m^n$.

Using the result given in Exercise 1.6, we get

$$\text{var}\,(X_{n+1}) = E(X_n)\,\text{var}\,(\zeta_r) + [E(\zeta_r)]^2\,\text{var}\,(X_n) \tag{2.9}$$

$$= m^n\sigma^2 + m^2\,\text{var}\,(X_n) \tag{2.9a}$$

We can find var (X_n) from (2.9a) either by induction or by solving the non-homogeneous difference equation. We employ the latter method.

A particular solution of the difference equation (2.9a) is given by

$$\text{var}\,(X_n) = \frac{\sigma^2 m^n}{m - m^2}$$

and a general solution of the homogeneous equation corresponding to (2.9a) is given by var $(X_n) = A(m^2)^n$, where A is a constant. Thus the complete solution of (2.9a) is

$$\text{var}\,(X_n) = A(m^2)^n + \frac{\sigma^2 m^n}{m - m^2}. \tag{2.10}$$

Noting that var $(X_1) = \sigma^2$, we get $A = \sigma^2/[m\,(m-1)]$ so that (2.10) yields

$$\text{var}\,(X_n) = \frac{m^{n-1}\,(m^n - 1)}{(m-1)}\,\sigma^2,\ n = 1, 2, \ldots.$$

The result holds for all n and $m \neq 1$. By taking limit as $m \to 1$, one gets the corresponding result for $m = 1$.

When $m = 1$, $\quad$ var $(X_n) = n\sigma^2$.

When $m = 1$, the variance of X_n increases linearly and when $m > 1$ $(m < 1)$ it increases (decreases) geometrically with n.

According as $m < 1, = 1$, or > 1, the Galton-Watson process is referred to as *subcritical, critical* or *supercritical* respectively.

We now come to the problem originally posed by Galton.

9.3 Probability of Extinction

Definition. By extinction of the process it is meant that the random sequence $\{X_n\}$ consists of zeros for all except a finite number of values of n. In other words, extinction occurs when Pr $\{X_n = 0\}$, for some value of n. Clearly, if $X_n = 0$ for $n = m$, then $X_n = 0$ for $n > m$; also $\Pr\{X_{n+1} = 0 \mid X_n = 0\} = 1$.

Theorem 9.3. If $m \leqslant 1$, the probability of ultimate extinction is 1. If $m > 1$, the probability of ultimate extinction is the positive root less than unity of the equation

$$s = P(s). \tag{3.1}$$

Proof: Let $q_n = \Pr\{X_n = 0\}$, i.e. q_n is the probability that extinction occurs at or before the nth generation. Clearly $q_n = P_n(0)$, $q_1 = P_1(0) = P(0) = p_0$ and from (2.5)

$$q_n = P(q_{n-1}). \tag{3.2}$$

If $p_0 = 0$, then $q_1 = 0$, $q_2 = 0$, $\ldots$, i.e. if the probability of no offspring is zero, extinction can never occur. If $p_0 = 1$, then $q_1 = 1$, $q_2 = 1$, $\ldots$, i.e. if the probability of no offspring is one then the extinction is certain to occur right after the 0th generation. So we confine ourselves to the case $0 < p_0 < 1$.

As $P(s)$ is a strictly increasing function of s, $q_2 = P(q_1) = P(p_0) > P(0) = q_1$. Assuming that $q_n > q_{n-1}$, we get $q_{n+1} = P(q_n) > P(q_{n-1}) = q_n$ and by induction $q_1 < q_2 < q_3 \ldots$. The monotone increasing sequence $\{q_n\}$ is bounded above by 1. Hence q_n must have a limit $\lim_{n\to\infty} q_n = q$ (say), $0 \leqslant q \leqslant 1$; q is the probability of ultimate extinction. From (3.2) it follows that q satisfies $q = P(q)$, i.e. q is a root of the equation (3.1),

$$s = P(s).$$

We now investigate further about the root. First, we show that q is the smallest positive root of (3.1). Let s_0 be an arbitrary positive root of (3.1). Then $q_1 = P(0) < P(s_0) = s_0$ and assuming $q_m < s_0$, we get $q_{m+1} = P(q_m) < P(s_0) = s_0$ and by induction $q_n < s_0$ for all n. Thus $q = \lim_{n\to\infty} q_n \leqslant s_0$, which implies that q is the smallest positive root of (3.2).

For this, we consider the graph of $y = P(s)$ in $0 \leqslant s \leqslant 1$; it starts with the point $(0, p_0)$ and ends with the point $(1, 1)$; the curve lying entirely in the first quadrant, is convex as $P(s)$ is an increasing function. So the curve $y = P(s)$ can intersect the line $y = s$ in at most two points, one of which is the end point $(1, 1)$, i.e. the equation (3.1) has at most two roots, one of which is unity. Two cases are now to be considered (see Figs. 9.1 and 9.2).

Case I. The curve $y = P(s)$ lies entirely above the line $y = s$; in which case $(1, 1)$ is the only point of intersection, i.e. unity is the unique root of $s = P(s)$ so that $q = \lim_{n\to\infty} q_n = 1$. Then

$$P(1) - P(s) = 1 - P(s) \leqslant 1 - s,$$

so that $\lim\limits_{s\to\infty} \dfrac{P(1) - P(s)}{1 - s} \leqslant 1$, i.e. $P'(1) \leqslant 1$.

Thus $\lim\limits_{n\to\infty} q_n = 1$ when $P'(1) = m \leqslant 1$.

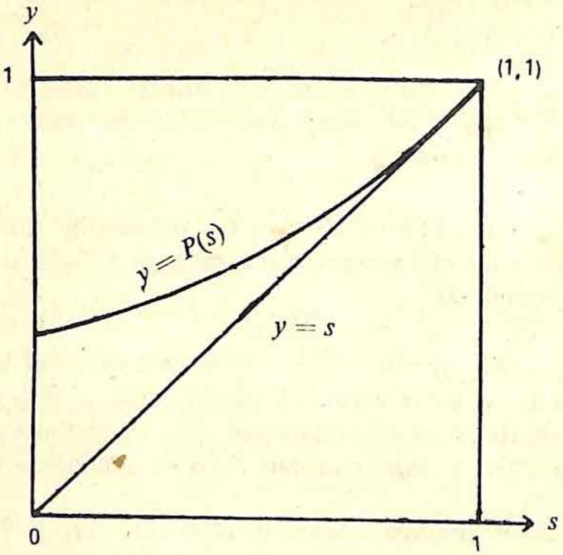


Fig. 9.1 $(m \leqslant 1)$

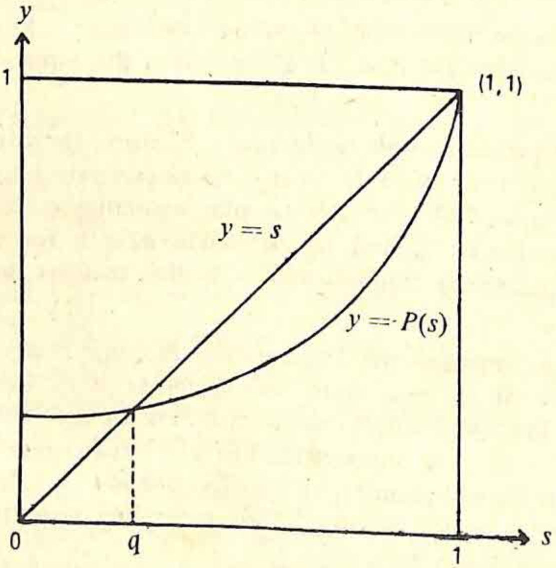


Fig. 9.2 $(m > 1)$

Case II. The curve $y = P(s)$ intersects $y = s$ at another point $(\delta, P(\delta))$ such that $\delta = P(\delta)$, $\delta < 1$, i.e. there is another root of (3.1), $\delta < 1$; the curve $y = P(s)$, being convex, lies below the line $y = s$ in $(\delta, 1)$, and above $y = s$ in $(0, \delta)$, i.e. $P(s) < s$ in $\delta < s < 1$ and $P(s) > s$ in $0 < s < \delta$. Then $q_1 = P(0) < P(\delta) = \delta$ and assuming $q_m < \delta$, we get $q_{m+1} = P(q_m) < P(\delta) = \delta$ and by induction $q_n < \delta$ for all n. Hence $\lim\limits_{n\to\infty} q_n = \delta$ so that $q = \delta < 1$.

Now by the mean value theorem considered in the interval $[\delta, 1]$, there is a value ξ in $\delta < \xi < 1$ such that $P'(\xi) = \frac{P(1) - P(\delta)}{1 - \delta} = 1$, and as the derivative is monotone $P'(1) > 1$. Thus we find that q is the root less than unity of (3.1) when $m = P'(1) > 1$.

We have thus proved the theorem completely. ▲

Note: That q is a root of $s = P(s)$ can also be seen by conditioning on the number of offsprings of the first generation. We have

$$\begin{aligned} q &= \Pr\{\text{ultimate extinction}\} \\ &= \sum_{k=0}^{\infty} \Pr\{\text{ultimate extinction} \mid X_1 = k\}. \Pr\{X_1 = k\}. \end{aligned}$$

Given that $X_1 = k$, the population will be extinct *iff* each of the k families started by members of the 1st-generation becomes extinct. It is assumed that families behave independently, hence

$$\Pr\{\text{ultimate extinction} \mid X_1 = k\} = q^k.$$

Thus
$$q = \sum_{k=0}^{\infty} q^k p_k = P(q).$$

Theorem 9.4. Whatever be the value of $E(X_1) = m$, we have, as $n \to \infty$, $\lim \Pr\{X_n = 0\} = q$ and $\lim \Pr\{X_n = k\} = 0$, for any finite positive integral k.

Proof: We first show that $\lim_{n\to\infty} P_n(s) = q$, from which the above result will follow.

Consider the case $m \leqslant 1$, when $P(s) = s$ has the unique root $q = 1$. In $0 \leqslant s \leqslant q$, $P(s) \leqslant P(q) = q$, and $P_2(s) \leqslant P_2(q) = P(P_1(P)) = P(q) = q$. Assuming that $P_m(s) \leqslant q$, we get $P_{m+1}(s) \leqslant q$ and by induction $P_n(s) \leqslant q$ for all n. Again $P_n(s) \geqslant P_n(0) = q_n$; thus $q_n \leqslant P_n(s) \leqslant q$.

Hence
$$\lim_{n\to\infty} P_n(s) = q, \quad 0 \leqslant s \leqslant q.$$

Consider the case $m > 1$, when q is the root less than 1 of $P(x) = x$. In $q < s < 1$, the curve $y = P(x)$ then lies below the line $y = x$, and $q < P(s) < s < 1$. Again $P_2(s) = P(P_1(s)) = P(P(s)) > P(q) = q$. Assuming that $P_m(s) > q$, we get $P_{m+1}(s) > q$, so that, by induction $P_n(s) > q$ for all n. Again $P_2(s) = P_1(P(s)) < P_1(s)$ and assuming that $P_m(s) < P_{m-1}(s)$, we get $P_n(s) < P_{n-1}(s)$ for all n.

Thus in
$$q < s < 1,$$
$$q < P_n(s) < P_{n-1}(s) < \ldots$$
so that
$$\lim_{n\to\infty} P_n(s) \geqslant q.$$

Suppose, if possible, that $\lim_{n\to\infty} p_n(s) = \alpha > q$, then $P(\alpha) < \alpha$, and $\lim_{n\to\infty} P_{n+1}(s) = \lim_{n+\infty} P(P_n(s))$, and we get a contradiction which is due to our supposition that $\alpha > q$. Thus

$$\lim_{n\to\infty} P_n(s) = q.$$

So, whatever be the value of $E(X_1) = m$, $\lim_{n\to\infty} P_n(s) = q$ is independent of s (for all $s < 1$). In other words, for all $s < 1$,

$$\lim_{n\to\infty} \sum_{k=0}^{\infty} \Pr\{X_n = k\}\, s^k = q.$$

This implies that the coefficients of s^k for $k \geqslant 1$ (except, possibly, for infinitely large k) all tend to 0, while the constant term tends to q. Thus, as $n \to \infty$

$$\Pr\{X_n = k\} \to 0, \text{ for finite positive integral } k,$$

and $\Pr\{X_n = 0\} \to q.$

Since $P_n(1) = 1$, we have, as $n \to \infty$

$$\Pr\{X_n \to \infty\} \to 1 - q. \tag{3.3}$$

Note: The above result also follows from the general theory of Markov chains applied to the chain $\{X_n\}$, for which each of the states $k = 1, 2, 3, \ldots$ is transient while the state 0 is absorbing. We shall now consider another interesting result.

Theorem 9.5. We have, for $r, n = 0, 1, 2, \ldots$

$$E\{X_{n+r} \mid X_n\} = X_n m^r. \tag{3.4}$$

Proof: For $r = 1$, and $n = 0, 1, 2, \ldots$, we have

$$E\{X_{n+1} \mid X_n\} = E\{\sum_{k=1}^{X_n} \zeta_k \mid X_n\} = \sum_{k=1}^{X_n} E\{\zeta_k\}$$

$$= mX_n.$$

Assume that (3.4) holds for $r = k$; then $E\{X_{n+k} \mid X_n\} = X_n m^k$. Noting the Markov nature of X_n, we get

$$\begin{aligned} E\{X_{n+k+1} \mid X_n\} &= E[E\{X_{n+k+1} \mid X_{n+k}, X_{n+k-1}, \ldots, X_n\} \mid X_n] \\ &= E[E\{X_{n+k+1} \mid X_{n+k}\} \mid X_n] \\ &= E\{mX_{n+k} \mid X_n\} = m\{X_n m^k\} \\ &= X_n m^{k+1}, \end{aligned}$$

so that the result holds for $r = k + 1$.

Thus by induction, we have the result. ▲

Note: Another variable of interest is $W_n = X_n/m^n$, $n = 0, 1, 2, \ldots$. $\{W_n\}$ forms a Markov chain. We have $E\{W_n\} = 1$ and for $m > 1$,

$$E\{W_n^2\} = \frac{1}{m^{2n}} E(X_n^2) = \frac{1}{m^{2n}} \left\{ m^{2n} + \frac{m^{n-1}(m^n - 1)\sigma^2}{(m-1)} \right\}$$

$$= 1 + \frac{\sigma^2}{m^2 - m}(1 - m^{-n}).$$

Dividing both sides of (3.4) by m^{n+r}, we get

$$E\{W_{n+r} \mid W_n\} = W_n \tag{3.5}$$

and since $\{W_n\}$ is also a Markov chain,

$$E\{W_{n+r} \mid W_n, W_{n-1}, \ldots, W_0\} = E\{W_{n+r} \mid W_n\} = W_n. \tag{3.6}$$

For limiting distribution of W_n and its use in the study of the distribution of X_n for large n and for other results see Harris (1963).

We now consider some simple examples; to fix the ideas numerical values have been taken.

Example 3(a). Let p_k, $k = 0, 1, 2$ be the probability that an individual in a generation generates k offsprings. Then $P(s) = p_0 + p_1 s + p_2 s^2$, and $P_2(s)$, $P_3(s)$ can be calculated by simple algebra. The probability of extinction is one if $m \leqslant 1$; if $m > 1$, it is given by the root less than 1 of $s = P(s)$. Suppose $p_0 = 2/3, p_1 = 1/6, p_2 = 1/6$; then $m = 1/2 < 1$. The equation $s = P(s)$ becomes $s^2 - 5s + 4 = 0$ with roots 1 and 4; the probability of extinction is 1. Suppose $p_0 = 1/4$, $p_1 = 1/4$, $p_2 = 1/2$; then $m = 5/4 > 1$; the equation $s = P(s)$ has the roots $1/2$ and 1. The root $1/2$ smaller than unity gives the probability of extinction. Note that the probability of extinction is p_0/p_2 or 1 according as $p_0 < p_2$ or $p_0 \geqslant p_2$ and also that $p_0 <$ (or $\geqslant$) p_2 iff $m >$ (or $\leqslant$) 1.

Example 3(b). Let the probability distribution of the number of offsprings generated by an individual in a generation be Poisson with mean λ, i.e. $P(s) = e^{\lambda(s-1)}$. It can be easily seen that the graph of $P(s)$ in $0 \leqslant s \leqslant 1$ (i.e. between the points $(0, e^{-\lambda})$ and $(1, 1)$) is convex, and that the curve of $y = P(s)$ always lies above $y = s$ when $\lambda \leqslant 1$, there being no other root of $s = P(s)$ except unity in $(0, 1)$; the probability of extinction is then 1. When $\lambda > 1$, the curve $y = P(s)$ intersects $y = s$ in another point whose s-coordinate has a value < 1 and the probability of extinction will be this value of s. For example, if $\lambda = 2$, it can be seen that $s = e^{2(s-1)}$ has a root approximately equal to 0.2 which is smaller than 1, and the probability of extinction is $q \sim 0.2$.

Example 3(c). Let the distribution of the number of offsprings be geometric with $p_k = b(1-b)^k$, $k = 0, 1, 2, \ldots$ $(0 < b < 1)$. Then $m = (1-b)/b$ and $P(s) = b/(1 - s(1-b))$. The equation $s = P(s)$ has the roots 1 and $b/(1-b)$. If $m \leqslant 1$, then the probability of extinction is 1; if $m > 1$, the root $b/(1-b) < 1$, and the probability of extinction is equal to the root $b/(1-b)$.

Example 3(d). Let $p_k = b\,c^{k-1}$, $k = 1, 2, \ldots$, $0 < b, c$, $b + c < 1$ and $p_0 = 1 - \sum_{k=1}^{\infty} p_k$. Then $m = b/(1-c)^2$. We have

$$P(s) = 1 - \frac{b}{1-c} + \frac{bs}{1-cs}. \tag{3.7}$$

The quadratic equation $s = P(s)$ has the roots

$$1 \text{ and } \frac{1-(b+c)}{c(1-c)} = s_0 \text{ (say).}$$

If $m = 1$, then $s_0 = 1$ and the probability of extinction is 1; if $m > 1$, $s_0 < 1$, and the probability of extinction is $q = s_0 (< 1)$.

This model was applied in a series of interesting papers by Lotka to find out the probability of extinction for American male lines of descent. The values estimated by him (in 1939) from census figures of 1920 give $b = 0.2126$, $c = 0.5893$ $(m \sim 1.25 > 1)$ and the probability of extinction $q = s_0 = 0.819$.

Note: It is not always possible to put the generating functions $P_n(s)$ in closed form. The generating functions $P(s)$ obtained in examples 3(c) and 3(d) are of an interesting form; they may be considered as particular cases of the more general fractional linear form (or general bilinear form)

$$P(s) = \frac{\alpha + \beta s}{\gamma + \delta s}, \ \alpha\delta - \beta\gamma \neq 0.$$

When $P(s)$ is of the above form, $P_n(s)$ is also of the same form

$$P_n(s) = \frac{\alpha_n + \beta_n s}{\gamma_n + \delta_n s}$$

where $\alpha_n, \beta_n, \gamma_n, \delta_n$ are functions of $\alpha, \beta, \gamma, \delta$.

Further, it may be noted that the equation $s = P(s)$ (where $P(s)$ is of fractional linear form) has two finite solutions 1 and s_0 and that $s_0 <, =, >$ or 1 according as $m = P'(1) >, =,$ or < 1.

Example 3(e). $P_n(s)$ for Lotka's model considered in Example 3(d) above. For any two points u, v, we get

$$\frac{P(s) - P(u)}{P(s) - P(v)} = \frac{s - u}{s - v}\frac{1 - cv}{1 - cu}.$$

Put $u = s_0$, $v = 1$, then $P(s_0) = s_0$, $P(v) = 1$, so that

$$\frac{P(s) - s_0}{P(s) - 1} = \frac{s - s_0}{s - 1}\ \frac{1 - c}{1 - cs_0}$$

whence

$$\frac{1 - c}{1 - cs_0} = \left\{\frac{P(s) - s_0}{s - s_0}\right\} \Big/ \left\{\frac{P(s) - 1}{(s - 1)}\right\}.$$

Let $m \neq 1$, then taking limits of the right hand side as $s \to 1$, we get

$$\frac{1 - c}{1 - cs_0} = \frac{1}{m}.$$

Hence

$$\frac{P(s) - s_0}{P(s) - 1} = \left(\frac{1}{m}\right)\frac{s - s_0}{s - 1}.$$

Thus

$$\frac{P_2(s) - s_0}{P_2(s) - 1} = \frac{P(P(s)) - s_0}{P(P(s)) - 1} = \left(\frac{1}{m}\right)\frac{P(s) - s_0}{P(s) - 1}$$

$$= \left(\frac{1}{m^2}\right)\frac{s - s_0}{s - 1}$$

and on iteration

$$\frac{P_n(s)-s_0}{P_n(s)-1} = \left(\frac{1}{m^n}\right)\frac{s-s_0}{s-1}, \; n = 1, 2, \ldots.$$

Solving for $P_n(s)$, we get

$$P_n(s) = 1 - m^n\left(\frac{1-s_0}{m^n-s_0}\right) + \frac{m^n\left(\dfrac{1-s_0}{m^n-s_0}\right)^2 s}{1-\left(\dfrac{m^n-1}{m^n-s_0}\right)s}, \; m \neq 1. \quad (3.8)$$

If $m = 1$, then $s_0 = 1$, and $P(s) = \dfrac{c+(1-2c)s}{1-cs}$

and
$$P_n(s) = \frac{nc+(1-c-nc)s}{(1-c+nc)-ncs}. \quad (3.9)$$

Limiting Results

Suppose that $m = 1$, then $\sigma^2 = 2c/(1-c)$,

$$n\Pr\{X_n > 0\} = n\{1-P_n(0)\} = \frac{n(1-c)}{1-c+nc}$$

and
$$\lim_{n\to\infty} n\Pr\{X_n > 0\} = \frac{1-c}{c} = \frac{2}{\sigma^2} \quad \text{(see Theorem 9.8(a))}.$$

Suppose that $m < 1$, then $s_0 > 1$ and

$$\lim_{n\to\infty} m^{-n}\Pr\{X_n > 0\} = \lim_{n\to\infty} m^{-n}\{1-P_n(0)\}$$
$$= \lim_{n\to\infty}\frac{1-s_0}{m^n-s_0} = \frac{s_0-1}{s_0}.$$

Again
$$\sum_k \Pr\{X_n = k \mid X_n > 0\}s^k = \frac{P_n(s)-P_n(0)}{1-P_n(0)}$$
$$= 1 - \frac{1-P_n(s)}{1-P_n(0)}$$
$$= \frac{\left(\dfrac{1-s_0}{m^n-s_0}\right)s}{1-\left(\dfrac{m^n-1}{m^n-s_0}\right)s}.$$

Thus
$$\lim_{n\to\infty}\sum_k \Pr\{X_n = k \mid X_n > 0\}s^k = s\left(1-\frac{1}{s_0}\right)\left(1-\frac{s}{s_0}\right)^{-1}$$

and
$$\lim_{n\to\infty}\Pr\{X_n = k \mid X_n > 0\} = \left(1-\frac{1}{s_0}\right)\left(\frac{1}{s_0}\right)^{k-1}, \; k = 1, 2, \ldots.$$

In other words, for large n, the distribution of $\{X_n\}$, given $X_n > 0$ is geometric with mean $s_0/(s_0-1)$ and p.g.f.

$$b(s) = \frac{(1-1/s_0)s}{1-s/s_0}.$$

It can be easily verified that $b(s)$ satisfies the equation

$$b(P(s)) = mb(s) + 1 - m \quad \text{(see Theorem 9.9)}. \quad ▲$$

9.4 Distribution of Total Number of Progeny

Let X_n denote the size of nth generation, $n = 0, 1, 2, \ldots, X_0 = 1$. Then the random variable

$$Y_n = \sum_{k=0}^{n} X_k = 1 + X_1 + \ldots + X_n \tag{4.1}$$

denotes the total number of progeny, i.e. the number of descendants upto and including the nth generation and also including the ancestor.

Theorem 9.6. The p.g.f. $R_n(s)$ of Y_n satisfies the recurrence relation

$$R_n(s) = sP(R_{n-1}(s)), \tag{4.2}$$

$P(s)$ being the p.g.f. of the offspring distribution.

Proof: Let $Z_n = X_1 + \ldots + X_n$ and $G_n(s)$ be its p.g.f. Then $R_n(s) = sG_n(s)$. We have

$$G_n(s) = \sum_{k=0}^{\infty} \Pr\{Z_n = k\} s^k.$$

Now by conditioning on the size X_1 of the 1st generation, we get $\Pr\{Z_n = k\} = \sum_{i=0}^{\infty} \Pr\{$total number of descendants in the succeeding $(n-1)$ generations following the 1st is $k - i \mid X_1 = i\} \cdot \Pr\{X_1 = i\}$.

If the process starts with one ancestor then the probability of having r descendants in succeeding m generations is the coefficient of s^r in $G_m(s)$; and if it starts with i ancestors then the probability of having r descendants in the succeeding m generations will be the coefficient of s^r in $[G_m(s)]^i$. Thus

$$\begin{aligned}
\Pr\{Z_n = k\} &= \sum_{i=0}^{\infty} [\text{coefficient of } s^{k-i} \text{ in } \{G_{n-1}(s)\}^i]\, p_i \\
&= \sum_{i=0}^{\infty} p_i [\text{coefficient of } s^k \text{ in } \{sG_{n-1}(s)\}^i] \\
&= \text{coefficient of } s^k \text{ in } \sum_{i=0}^{\infty} p_i \{sG_{n-1}(s)\}^i \\
&= \text{coefficient of } s^k \text{ in } P(sG_{n-1}(s)).
\end{aligned}$$

Thus $$G_n(s) = \sum_{k=0}^{\infty} \Pr\{Z_n = k\} s^k = P(sG_{n-1}(s)),$$

whence $$R_n(s) = sG_n(s) = sP(R_{n-1}(s)).$$

Hence the theorem. ▲

From the recurrence relation (4.2) it is theoretically possible to calculate $R_1(s), R_2(s), \ldots$. We are however interested in the asymptotic behaviour of $R_n(s)$ for large n.

Theorem 9.7. We have

$$\lim_{n \to \infty} R_n(s) = H(s),$$

where $H(s) = \sum_{k=0}^{\infty} \rho_k s^k$ is the generating function of a sequence of non-negative numbers ρ_k.

The function $H(s)$ satisfies the relation

$$H(s) = sP(H(s)), \quad 0 < s < 1; \tag{4.3}$$

further $H(s)$ is the unique root of the equation

$$t = s\,P(t) \tag{4.4}$$

such that $H(s) \leqslant \xi$, where ξ is the smallest positive root of $x = P(x)$ and that $H(1) = \Sigma \rho_k = \xi$.

Proof: We have, for $0 < s < 1$,

$$R_2(s) = sP(R_1(s)) < sP(s) = R_1(s)$$

and assuming that $R_m < R_{m-1}(s)$, we get

$$R_{m+1}(s) = sP(R_m(s)) < sP(R_{m-1}(s)) = R_m(s)$$

and hence by induction $R_n(s) < R_{n-1}(s)$ for all $n > 0$. Thus for $s < 1$, $\{R_n(s)\}$ is a monotone decreasing sequence bounded below. Hence $\lim_{n\to\infty} R_n(s) = H(s)$ exists. From the continuity theorem of p.g.f.'s, it follows that $H(s)$, being the limit of a sequence of p.g.f.'s, is the generating function of a sequence of non-negative numbers ρ_k such that $H(1) = \Sigma \rho_k \leqslant 1$.

Taking limit of (4.2), we get

$$H(s) = sP(H(s)), \qquad 0 < s < 1,$$

i.e. for some fixed s (in $0 < s < 1$), $H(s)$ is a root of the equation

$$t = sP(t).$$

For fixed $s < 1$, $y = sP(t)$ is a convex function of t and the graph of $y = sP(t)$ intersects the line $y = t$ in at most two points. Let ξ be the smallest positive root of $x = P(x)$; clearly $\xi \leqslant 1$. The function $t - sP(t)$ is negative for $t = 0$ and positive for $t = \xi$, and remains positive for values of t between ξ and 1. Thus $t = sP(t)$ has exactly one root between 0 and ξ and no root between ξ and 1. The unique root of $t = sP(t)$ equals $H(s)$ and thus $H(s) < \xi$. Clearly $H(1)$ is a root of $x = P(x)$ and since ξ is the smallest root of this equation, $H(1) = \xi$. Thus the theorem is completely established. ▲

Note: $H(1) = 1$ or < 1 depending on whether $1 + X_1 + X_2 + \dots$ is finite or not with probability one; and $H(1) = 1$ whenever $m \leqslant 1$.

9.5 Conditional Limit Laws

A. Critical Processes

In Example 3(e) we obtained some limiting results. We shall obtain here some general results. Consider a critical (i.e. with $m = 1$) G.W. process. The probability of extinction is 1. Thus from Theorem 9.4, we get $\Pr\{X_n \to 0\} = 1$. We also have $\operatorname{var}(X_n) = n\sigma^2 \to \infty$. The distribution of X_n given that $X_n > 0$ is of considerable interest.

Lemma. For a G.W. process with $m = 1$ and $\sigma^2 < \infty$, we have

$$\lim_{n\to\infty} \frac{1}{n}\left\{\frac{1}{1-P_n(s)} - \frac{1}{1-s}\right\} \to \frac{\sigma^2}{2} \tag{5.1}$$

uniformly in $0 \leqslant s < 1$.

Proof: Let $0 \leqslant s < 1$ and $P'''(1) < \infty$. Using Taylor's expansion of $P(s)$ in the neighbourhood of 1, we get

$$P(s) = s + \frac{\sigma^2}{2}(1-s)^2 + r(s)(1-s)^2, \tag{5.2}$$

where

$$r(s) \to 0 \text{ as } s \to 1.$$

Thus

$$\begin{aligned}\frac{1}{1-P(s)} - \frac{1}{1-s} &= \frac{P(s)-s}{(1-s)(P(s)-s)}\\ &= \frac{1-s}{1-P(s)}\left\{\frac{\sigma^2}{2} + r(s)\right\}\\ &= \left\{\frac{\sigma^2}{2} + r(s)\right\}\left[1-(1-s)\left\{\frac{\sigma^2}{2} + r(s)\right\}\right]^{-1}\\ &= \frac{\sigma^2}{2} + R(s),\end{aligned} \tag{5.3}$$

where $R(s) \to 0$ as $s \to 1$, and R is bounded. Again using (4.6), we get

$$P_2(s) = P(P(s)) = P(s) + \frac{\sigma^2}{2}(1-P(s))^2 + r(P(s))(1-P(s))^2$$

so that

$$\frac{1}{1-P_2(s)} - \frac{1}{1-P(s)} = \frac{\sigma^2}{2} + R(P(s)) \tag{5.4}$$

and

$$\frac{1}{2}\left\{\frac{1}{1-P_2(s)} - \frac{1}{1-s}\right\} = \frac{\sigma^2}{2} + \frac{1}{2}\{R(s) + R(P(s))\}.$$

Iterating one gets

$$\frac{1}{n}\left\{\frac{1}{1-P_n(s)} - \frac{1}{1-s}\right\} = \frac{\sigma^2}{2} + \frac{1}{n}\sum_{k=0}^{n-1} R(P_k(s)).$$

Since $P_n(0) \leqslant P_n(s) < 1$ and $P_n(0) \to 1$ from the left, the convergence of $P_n(s) \to 1$ is uniform. Hence the lemma. ▲

We shall now use the lemma to establish the following interesting limit laws (of which (a) is due to Kolmogorov and (b), (c) are due to Yaglom).

Theorem 9.8. If

$$m = 1, \sigma^2 < \infty,$$

then

(a)
$$\lim_{n\to\infty} n \Pr\{X_n > 0\} = \frac{2}{\sigma^2}$$

(b) $$\lim_{n\to\infty} E\left\{\frac{X_n}{n}\,\middle|\, X_n > 0\right\} = \frac{\sigma^2}{2}$$

(c) $$\lim_{n\to\infty} \Pr\left\{\frac{X_n}{n} > u \mid X_n > 0\right\} = \exp\left(-\frac{2u}{\sigma^2}\right),\ u \geqslant 0.$$

Proof: (a) $n \Pr\{X_n > 0\} = n\{1 - P_n(0)\}$

$$= \left[\frac{1}{n}\left\{\frac{1}{1-P_n(0)} - 1\right\} + \frac{1}{n}\right]^{-1}.$$

Thus from the lemma (taking $s = 0$), we get

$$\lim_{n\to\infty} n \Pr\{X_n > 0\} = \lim_{n\to\infty}\left[\frac{\sigma^2}{2} + \frac{1}{n}\right]^{-1} = \frac{2}{\sigma^2}\cdot$$

(b) We have

$$1 = E\{X_n\} = E\{X_n \mid X_n > 0\}.\Pr\{X_n > 0\} + 0.\Pr(X_n = 0\}$$

so that

$$E\{X_n \mid X_n > 0\} = \frac{1}{\Pr\{X_n > 0\}} = \frac{n\sigma^2}{2}, \text{ (from (a))}.$$

Thus $$\lim_{n\to\infty} E\left\{\frac{X_n}{n}\,\middle|\, X_n > 0\right\} = \frac{\sigma^2}{2}.$$

(c) Let $u > 0$, and

$$dF(u) = \Pr\left\{u < \frac{X_n}{n} < u + du \mid X_n > 0\right\}; \tag{5.5}$$

then taking L.T. (see equation (3.1a) Ch. 1), we get

$$\int_0^\infty \exp(-\alpha u)\, dF(u) = E\left\{\exp\left(-\frac{X_n}{n}\right)\middle|\, X_n > 0\right\}. \tag{5.6}$$

Now $$E\left\{\exp\left(-\frac{X_n}{n}\right)\right\} = E\left\{\exp\left(-\frac{X_n}{n}\right)\middle|\, X_n > 0\right\}.\Pr\{X_n > 0\} + 1.\Pr\{X_n = 0\}$$

and since $P_n(s) = E\{s^{X_n}\}$ is the p.g.f. of X_n, we get

$$P_n\left(\exp\left(-\frac{\alpha}{n}\right)\right) = E\left\{\exp\left(-\frac{\alpha X_n}{n}\right)\middle|\, X_n > 0\right\}\{1 - P_n(0)\} + P_n(0).$$

Thus

$$E\left\{\exp\left(-\frac{\alpha X_n}{n}\right)\middle|\, X_n > 0\right\} = \frac{P_n(\exp(-\alpha/n)) - P_n(0)}{1 - P_n(0)}$$

$$= 1 - \frac{1 - P_n(\exp(-\alpha/n))}{1 - P_n(0)}. \tag{5.7}$$

Now as $n \to \infty$

$$n\left\{1 - P_n\left(\exp\left(-\frac{\alpha}{n}\right)\right)\right\} \to \frac{2}{\sigma^2} \text{ (from (a))}$$

and from the basic lemma (because of uniform convergence), we get

$$\frac{1}{n\{1 - P_n(\exp(-\alpha/n)\}} = \frac{1}{n}\left\{\frac{1}{1 - P_n(\exp(-\alpha/n))} - \frac{1}{1 - \exp(-\alpha/n)}\right\}$$
$$+ \frac{1/n}{1 - \exp(-\alpha/n)} \to \frac{\sigma^2}{2} + \frac{1}{\alpha}.$$

Thus from (5.6) and (5.7), we get, as $n \to \infty$

$$\int_0^\infty \exp(-\alpha u)\, dF(u) \to 1 - \frac{\sigma^2/2}{\sigma^2/2 + 1/\alpha}$$
$$= \frac{1}{1 + \alpha\sigma^2/2}.$$

Since L.T. of $\frac{2}{\sigma^2} \exp\left(-\frac{2u}{\sigma^2}\right)$ is $\frac{1}{1 + \alpha\sigma^2/2}$,

we have

$$\lim_{n\to\infty} \Pr\left\{u < \frac{X_n}{n} < u + du \mid X_n > 0\right\} = \frac{2}{\sigma^2} \exp\left(-\frac{2u}{\sigma^2}\right),$$

which establishes the exponential limit law. ▲

B. Subcritical Processes

Theorem 9.9. (*Yaglom's theorem*). For a Galton-Watson process with $m < 1$,

$$\lim_{n\to\infty} \Pr\{X_n = j \mid X_n > 0\} = b_j,\ j = 1, 2, \ldots \tag{5.8}$$

exists, and $\{b_j\}$ gives a probability distribution whose p.g.f. $B(s) = \sum_{j=1}^{\infty} b_j s^j$ satisfies the equation

$$B(P(s)) = m\, B(s) + 1 - m \tag{5.9}$$

or,
$$1 - B(P(s)) = m(1 - B(s)). \tag{5.9a}$$

Further, $\sum_{j=1}^{\infty} jb_j = 1/\varphi(0)$, where $\varphi(0) = \lim_{n\to\infty} \Pr\{X_n > 0\}/m^n$.

Proof: Using Taylor's expansion around $s = 1$, we get

$$P(s) = 1 - m(1 - s) + (1 - s)\, r(s),\ 0 \leqslant s \leqslant 1 \tag{5.10}$$

or,
$$\frac{1 - P(s)}{1 - s} = m - r(s). \tag{5.10a}$$

Consider the function $r(s)$ in $0 \leqslant s < 1$; we have $r(0) = m - (1 - p_0) \geqslant 0$ and $\lim_{s\to 1-0} r(s) = 0$. Further as $P(s)$ is a convex function

$$P'(s) \leqslant \frac{1 - P(s)}{1 - s},$$

so that
$$r'(s) = (1 - s)^{-1}\left\{\frac{1 - P(s)}{1 - s} - P'(s)\right\} \leqslant 0.$$

Thus $r(s)$ is monotone decreasing, is bounded above by m and $r(s) \to 0$ as $s \to 1$. Replacing s by $P_{k-1}(s)$ in (5.10a), we get

$$\frac{1 - P_k(s)}{1 - P_{k-1}(s)} = m\,\{1 - r\,(P_{k-1}(s))/m\}. \tag{5.11}$$

Putting $k=1, 2, 3, \ldots, n$ and taking products of both sides, we get

$$\frac{1 - P_n(s)}{1 - s} = m^n \prod_{k=0}^{n-1} \{1 - r(P_k(s))/m\}. \tag{5.12}$$

Since $0 \leqslant r/m \leqslant 1$, the sequence $\left\{\dfrac{1 - P_n(s)}{m^n(1 - s)}\right\}$ is monotone decreasing in n, and we have

$$\lim_{n\to\infty} \frac{1 - P_n(s)}{m^n(1 - s)} = \varphi\,(s) \geqslant 0.$$

Putting $s = 0$, we get

$$\varphi\,(0) = \lim_{n\to\infty} \frac{1 - P_n(0)}{m^n} = \lim_{n\to\infty} \frac{\Pr\,\{X_n > 0\}}{m^n}. \tag{5.13}$$

Let $b_{jn} = \Pr\,\{X_n = i \mid X_n > 0\}$ and $B_n(s) = \sum_{j=1}^{\infty} b_{jn}\, s^j$ be the p.g.f. of $\{b_{jn}\}$. Then

$$B_n(s) = \frac{P_n(s) - P_n(0)}{1 - P_n(0)} = 1 - \frac{1 - P_n(s)}{1 - P_n(0)} \tag{5.14}$$

$$= 1 - (1 - s) \prod_{k=0}^{n-1} \frac{1 - r\,(P_k(s))/m}{1 - r\,(P_k(0))/m} \qquad \text{(from (5.12))}.$$

Since $P_k(s) \geqslant P_k(0)$, and $r(s)$ is monotone decreasing, $r(P_k(s)) \leqslant r(P_k(0))$, and so each factor of the product on the r.h.s. expression is larger than 1. Thus $B_n(s)$ is monotone decreasing and tends to a limit $B(s)$ as $n \to \infty$, i.e.

$$B(s) = \sum_{j=1}^{\infty} b_j\, s^j, \text{ where } b_j = \lim_{n\to\infty} b_{jn} = \lim_{n\to\infty} \Pr\,\{X_n = j \mid X_n > 0\}.$$

Clearly $B(0) = 0$.

$$\begin{aligned} B(P_k(0)) &= \lim_{n\to\infty} B_n(P_k(0)) \\ &= \lim_{n\to\infty} \left\{1 - \frac{1 - P_n(P_k(0))}{1 - P_n(0)}\right\} \\ &= 1 - \lim_{n\to\infty} \left\{\frac{1 - P_k(P_n(0))}{1 - P_n(0)}\right\} \\ &= 1 - \lim_{s\to 1} \frac{1 - P_k(s)}{1 - s} \\ &\qquad \text{(since } m < 1,\ P_n(0) \to 1 \text{ as } n \to \infty) \\ &= 1 - m^k \\ &\qquad \text{(taking limit of (5.12) as } s \to 1). \end{aligned}$$

It follows that lim $B(P_k(0)) \to 1$. As $m < 1$, $P_k(0) \to 1$ as $k \to \infty$.

Hence $B(s) \to 1$ as $s \to 1$. Thus $B(s)$ is the p.g.f. of $\{b_j\}$.

Further

$$\sum_{j=1}^{\infty} jb_j = B'(1) = \lim_{k\to\infty} \frac{1-B(P_k(0))}{1-P_k(0)}$$

$$= \lim_{k\to\infty} \frac{m^k}{1-P_k(0)}$$

$$= \lim_{k\to\infty} \frac{m^k}{\Pr\{X_k > 0\}} \to \frac{1}{\varphi(0)}.$$

Again, from (5.14)

$$B_n(P(s)) = 1 - \frac{1-P_n(P(s))}{1-P_n(0)}$$

$$= 1 - \frac{1-P_{n+1}(s)}{1-P_{n+1}(0)} \cdot \frac{1-P_{n+1}(0)}{1-P_n(0)} \tag{5.15}$$

$$\lim_{n\to\infty} \frac{1-P_{n+1}(s)}{1-P_{n+1}(0)} = \lim_{n\to\infty} (1-B_{n+1}(s)) = 1-B(s),$$

$$\lim_{n\to\infty} \frac{1-P_{n+1}(0)}{1-P_n(0)} = \lim_{s\to 1} \frac{1-P(s)}{1-s} = m.$$

Hence taking limits of both sides of (5.15), we get

$$B(P(s)) = 1-(1-B(s))m$$

$$= mB(s) + 1 - m.$$

Thus the theorem is proved. ▲

Remarks: The above simplified proof which does not involve moment restriction is due to Joffe (1967).

The equation (5.9a) is known as a modified *Schröder functional equation.* It is not always easy to obtain $B(s)$ from it for given $P(s)$. In Example 3(e) we obtained $B(s)$ directly.

The mean $\sum_{k=1}^{\infty} kb_k = 1/\varphi(0)$ of the limiting distribution is finite iff $E\{X_1 \log X_1\} = \sum_{k=1}^{\infty} p_k \{k \log k\} < \infty$ or $p_0 = 1$.

Before closing the discussion on G.W. processes we mention a few interesting innovations introduced in the basic process.

Generalisations of the Classical Galton-Watson Process

Branching processes with immigration: Theorem 9.4 states that for a G.W. process, $\Pr\{X_n \to 0\} = q$ and $\Pr\{X_n \to k\} = 0$, for finite k (and so $\Pr\{X_n \to \infty\} = 1-q$), q being the probability of extinction. Further $q=1$ for critical and subcritical processes. Thus left to themselves G.W. populations either die out or grow without limits. Immigration from outside into a critical or subcritical process could have stabilising effect on the population size. Apart from this aspect, immigration by itself is interesting from the point of view of theory and applications. Galton-Watson processes with immigration (briefly called G.W.I. processes) often arise in applications in such areas as traffic theory, statistical mechanics, genetics, neurophysiology etc.

Consider a G.W. process with offspring distribution $\{p_k\}$ (having p.g.f. $P(s)$ and mean $P'(1) = m$). (The process will be called underlying G.W. process.) Suppose that at time n, i.e. at the time of birth of nth generation there is an immigration of Y_n objects into the population, and that Y_n, $n = 0, 1, 2, \ldots$ are i.i.d random variables with p.g.f.

$$h(s) = \sum_{j=0}^{\infty} \Pr\{Y_n = j\}\, s^j = \Sigma\, h_j s^j,$$

i.e. with probability h_j, j immigrants enter the nth generation and contribute to the next generation in the same way as others already present do. The numbers of immigrants into successive generations are independent and all objects reproduce independently of each other and of the immigration process. The distribution $\{h_j\}$ will be called immigrant distribution. Let $a = h'(1)$ be the mean of this distribution.

Let $X_{(n)}$ be the number of objects at the nth generation and let

$$P_{(n)}(s) = \sum_{j=0}^{\infty} \Pr\{X_{(n)} = j\}\, s^j$$

be its p.g.f. The sequence $\{X_{(n)}, n = 0, 1, 2, \ldots\}$ defines a G.W.I process. The sequence is a Markov chain whose one-step transition probabilities are given by

$$p_{ij} = \text{coeff. of } s^j \text{ in } h(s)\,[P(s)]^i,\ i, j \in N.$$

Clearly,

$$P_{(n)}(s) = h(s)\, P_{(n-1)}(P(s)). \tag{5.16}$$

If $\lim_{n\to\infty} P_{(n)}(s) = F(s)$ exists, then one gets

$$F(s) = h(s)\, F(P(s)). \tag{5.17}$$

i.e. the limit, when it exists, satisfies the above functional equation. Now the question arises, when does the limit exist, and when does it define the p g.f. of a proper probability distribution or when does $\{X_{(n)}\}$ have a proper limit distribution?

Corresponding to Yaglom's result for subcritical G. W. processes, Heathcote (1965) obtained the following analogue for G.W.I. processes.

Consider a Galton-Watson process with immigration, having offspring distribution $\{p_j\}$ (with p.g.f. $P(s)$) and immigrant distribution $\{h_j\}$ (with p.g.f. $h(s)$). If $m < 1$, $a = h'(1) < \infty$, then, $\lim_{n\to\infty} \Pr\{X_{(n)} = j\} = d_j$, exists for $j \in N$, and the p.g.f. $F(s) = \sum_j d_j s^j$ satisfies the functional equation (5.17),

$$F(s) = h(s)\, F(P(s)).$$

Further,

$$F(1) = 1 \text{ iff } \sum_{j=1}^{\infty} h_j \log j < \infty.$$

It may be noted that, when

$$F(s) = \frac{1 - B(s)}{1 - s}$$

and $$h(s) = \frac{1-P(s)}{m(1-s)}$$

then the functional equation (eqn. (5.9) which appears in Yaglom's theorem)

$$B(P(s)) = mB(s) + 1 - m,$$

reduces to the functional equation (5.17). Thus, ordinary subcritical G.W. processes may always be regarded as forming a subclass of G.W.I. processes (at least as regards Yaglom's conditional limit theorem). The result for the critical process, due to Seneta (1970), may be stated as follows:

For a Galton-Watson process with immigration having $m = 1$, $\sigma^2 < \infty$, $a < \infty$, the random variable $X_{(n)}/n$ converges in distribution, as $n \to \infty$, to a random variable having gamma density,

$$\frac{u^{\alpha-1}\, e^{-u/\beta}}{\Gamma(\alpha)\, \beta^{\alpha}}, \quad u \geqslant 0,$$

where $$\alpha = 2a/\sigma^2, \ \beta = \sigma^2/2.$$

If $a = h'(1) = \sigma^2/2$, i.e. $\alpha = 1$, then the gamma density reduces to an exponential density, such that

$$\lim_{n\to\infty} \Pr\left\{\frac{X_{(n)}}{n} > u\right\} \to \exp\left\{-\frac{2u}{\sigma^2}\right\};$$

again from Theorem 9.8(c) we get

$$\lim_{n\to\infty} \left\{\frac{X_n}{n} > u \mid X_n > 0\right\} \to \exp\left(-\frac{2u}{\sigma^2}\right), \ u \geqslant 0,$$

for the G.W. process $\{X_n\}$ without immigration.

Thus in the critical case the effect of the conditioning of non-extinction $\{X_n > 0\}$ is the same as immigrating into the process at an average rate of $a = \sigma^2/2$.

Detailed studies on processes with immigration have been made by several authors (e.g. Heathcote, Seneta, Foster, Yang etc.). In Jagers' book, one can find treatment of some of them besides proof etc. of the above results.

Processes in Varying and Random Environments

It has been assumed that the offspring distribution $\{p_k\}$ remains the same in all the generations and that objects reproduce independently of others. It was Good who first pointed out the need to consider G.W. processes under varying environments. The assumption that the distribution $\{p_k\}$ remains the same in all the generations, was then replaced by the assumption that offspring distribution for the nth generation is of the form $\{p_{nk}\}$; several interesting results on processes in varying environments were obtained, among others, by Fearn (1979) and Jagers (1974) etc.

Smith and Wilkinson (1969) introduced yet another aspect. They postulated that the offspring distribution for each generation is *randomly* *chosen* from a class of all reproduction laws and the resulting processes are said to be in *random environments*. The topic has been pursued in great details in a number of papers by Wilkinson (1969), Athreya and Karlin (1970, 1971), Kaplan (1972) and others.

The models under these assumptions have wider scope for applications.

Multitype Galton-Watson Processes

A natural extension of Galton-Watson process is concerned with the case where the population consists of a finite number of types of objects. Such processes are known as *multitype Galton-Watson processes*. Mode (1971) deals exclusively and extensively with such processes, discusses several models for applications in various areas.

Suppose that a population of individuals (or objects) originates with a single ancestor and that there are k (finite) types of individuals. Let $p^{(i)}(r_1, r_2, \ldots, r_k)$ be the probability that an individual of type i produces r_j offsprings of type j, $j=1, 2, \ldots, k$. We introduce the vector notation:

$$\mathbf{r} = (r_1, r_2, \ldots, r_k),$$
$$\mathbf{s} = (s_1, s_2, \ldots, s_k),$$
$$\mathbf{e}_i = (\delta_{1i}, \delta_{2i}, \ldots, \delta_{ki}).$$

Let $f^{(i)}(\mathbf{s}) = \Sigma p^{(i)}(\mathbf{r}) s^{r_1} \ldots s^{r_k}$ be the p.g.f. of $\{p^{(i)}(\mathbf{r})\}$. Let

$$\mathbf{X}_n = (X_n^{(1)}, X_n^{(2)}, \ldots, X_n^{(k)})$$

represent the population size of k types in the nth generation.

Let m_{ij} be the expected number of offspring of type j produced by an object of type i. Then

$$m_{ij} = E\{X_1^{(j)} \mid \mathbf{X}_0 = \mathbf{e}_i\}$$
$$= \frac{\partial f^{(i)}}{\partial s_j}(1, 1, \ldots, 1),\ i, j = 1, 2, \ldots, k.$$

The matrix $M = (m_{ij})$ is the matrix of moments. Let $f_n^{(i)}(\mathbf{s})$ denote the p.g.f. of the number of objects in nth generation starting from one object of type i. A result analogous to (2.5) is as follows:

$$f_n^{(i)}(\mathbf{s}) = f^{(i)}(f_{n-1}^{(1)}(\mathbf{s}), f_{n-1}^{(2)}(\mathbf{s}), \ldots, f_{n-1}^{(k)}(\mathbf{s})),$$
$$f_0^{(i)}(\mathbf{s}) = s_i,\ i = 1, 2, \ldots, k;\ n = 0, 1, 2, \ldots.$$

The population becomes extinct when $\mathbf{X}_n = \mathbf{0}$ for some $n \geqslant 1$. Let

$$q_n^{(i)} = \Pr\{\mathbf{X}_n = \mathbf{0} \mid \mathbf{X}_0 = \mathbf{e}_i\}$$

and

$$q^{(i)} = \lim_{n\to\infty} q_n^{(i)};$$

i.e., $q^{(i)}$ denotes the probability of extinction given that the process started with one ancestor of type i. Let $\mathbf{q} = (q^{(1)}, q^{(2)}, \ldots, q^{(k)})$.

We confine ourselves to the case where the matrix of moments is of a certain type.

Let M be a $k \times k$ matrix with non-negative elements such that for some positive integer n, all the elements of M^n are strictly positive. Then there exists a positive eigenvalue ρ of M, which is greater than the absolute value of any other eigenvalue of M; ρ is also called the spectral radius of M.

When M is the matrix (m_{ij}), this eigenvalue ρ in the multitype G.W. case plays the role of the mean m of offspring distribution in the simple G.W. case. A multitype G.W, process is said to be subcritical, critical or supercritical depending on $\rho < 1, = 1$, or > 1. We now state, without proof, the *multitype analogue of Theorem 9.3*:

Let the matrix of moments M be positively regular and let ρ be its spectral radius.

(a) If $\rho \leqslant 1$, then $q_i = 1$, $i = 1, 2, \ldots, k$.

(b) If $\rho > 1$, then $q_i < 1$, $i = 1, 2, \ldots, k$.

(c) In either case, $\mathbf{q} = (q^{(1)}, q^{(2)}, \ldots, q^{(k)})$

is the smallest positive solution of the vector equation

$$\mathbf{q} = \mathbf{f}(\mathbf{q}).$$

i.e. $q^{(i)}$ is the smallest positive root of $q^{(i)} = f^{(i)}(\mathbf{q})$ for $i = 1, 2, \ldots, k$.

For proof and other results, the reader is referred to Athreya and Ney (1972) and Mode (1971).

9.6 Continuous Time Markov Branching Processes

So far we have been considering branching processes $\{X_n\}$ in discrete time; an object after one unit of time produces similar objects according to offspring distribution $\{p_k\}$. The lifetime of an object is one unit of time. Now we proceed to consider a generalisation such that the lifetimes of objects are i.i.d. random variables. Instead of the process $\{X_n, n = 0, 1, 2, \ldots\}$ we shall consider the process $\{X(t), t \geqslant 0\}$, where $X(t)$ equals the number of objects at time t. The process $\{X(t), t \geqslant 0\}$ may or may not be Markovian. In this section we shall confine to the Markovian case such that $\{X(t), t \geqslant 0\}$ will be a Markov process in continuous time with discrete state space. Such processes were considered in Sec. 4.5. We shall recall some of the basic results concerning such processes. We first note that the waiting times between changes of states (here lifetimes of objects) have independent exponential distributions. Let the process be time homogeneous.

Denote the transition probability by

$$p_{ij}(t) = \Pr\{t + t_1 = j \mid X(t_1) = i\},$$

and, in particular

$$p_{ij}(t) = \Pr\{X(t) = j \mid X(0) = i\}, j=0, 1, \ldots, t \geqslant 0.$$

The transition probabilities $p_{ij}(t)$ satisfy the Chapman-Kolmogorov equation

$$p_{ij}(t + T) = \sum_{k=0}^{\infty} p_{ik}(t)\, p_{kj}(T).$$

We shall assume here that $X(0) = 1$, i.e. the process starts with a single ancestor at time 0.

Let the probability that an object will split producing j objects during a small interval $(t, t+h)$ be denoted by $a_j h + o(h)$. The constants a_j (assumed to be independent of time) are the transition densities

$$a_j = \lim_{h\to 0} \frac{p_{1j}(h)}{h}, \, j \neq 1$$

$$a_1 = \lim_{h\to 0} \frac{p_{11}(h) - 1}{h}$$

so that

$$p_{1j}(h) = a_j h + o(h), \quad j = 0, 2, 3, \ldots$$

$$p_{11}(h) = 1 + a_1 h + o(h). \tag{6.1}$$

We have $a_j \geqslant 0, j = 0, 2, 3, \ldots$ and $\sum_{j=0}^{\infty} a_j = 0$.

Now objects act independently, and in one small interval, only one particle, on the average, will split, so that

$$\Pr\{X(t+h) = n+k-1 \mid X(t) = n\} = na_k h + o(h), \quad k = 0, 2, 3, \ldots$$

and $\Pr\{X(t+h) = n \mid X(t) = n\} = 1 + na_1 h + o(h)$.

Let $F(t, s)$ be the p.g.f. of $\{p_{1j}(t)\}$. We get

$$F(t, s) = \sum_{j=0}^{\infty} p_{1j}(t) s^j = \sum_{j=0}^{\infty} \Pr\{X(t) = j \mid X(0) = 1\} s^j \tag{6.2}$$

$$= E\{s^{X(t)}\},$$

and

$$F(0, s) = s. \tag{6.2a}$$

Definition. A Markov branching process $\{X(t), t \geqslant 0\}$ is a continuous time Markov chain having for state space the set of non-negative integers $N = \{0, 1, 2, \ldots\}$ and whose transition probabilities $p_{ij}(t)$ satisfy the relation

$$\sum_{j=0}^{\infty} p_{ij}(t) s^j = [\sum_{j=0}^{\infty} p_{1j}(t) s^j]^i = [F(t, s)]^i, \; i \geqslant 0, \; | s | \leqslant 1. \tag{6.3}$$

It is the property (6.3) that characterises and distinguishes a Markov branching process from other continuous time Markov chains with state space N (discussed in Sec. 4.5). The equation (6.3) is implied in the postulate that individuals act independently.

We shall now obtain a continuous time analogue of the relation (2.2) satisfied by a Galton-Watson process.

Theorem 9.10. We have

$$[F(t+T, s)]^i = [F(t, F(T, s))]^i, \; i \geqslant 0 \tag{6.4}$$

and in particular,

$$F(t+T, s) = F(t, F(T, s)). \tag{6.4a}$$

Proof: Using the relation (6.2) and the Chapman-Kolmogorov equation, we get

$$
\begin{aligned}
[F(t+T, s)]^i &= \sum_{j=0}^{\infty} p_{ij}(t+T)\, s^j,\ i \geqslant 0 \\
&= \sum_{j=0}^{\infty} s^j \{ \sum_{k=0}^{\infty} p_{ik}(t)\, p_{kj}(T) \} \\
&= \sum_{k=0}^{\infty} p_{ik}(t) \{ \sum_{j=0}^{\infty} p_{kj}(T)\, s^j \} \\
&= \sum_{k=0}^{\infty} p_{ik}(t) \{F(T, s)\}^k \\
&= [F(t, F(T, s))]^i,
\end{aligned}
$$

and, in particular,

$$F(t+T, s) = F(t, F(T, s))$$

We shall now proceed to obtain two differential equations (forward (6.6) and backward (6.7)), satisfied by $F(t, s)$; these can be obtained from the Chapman-Kolmogorov forward and backward differential equations respectively. We shall, however, use a slightly different method.

With the given initial conditions, the differential equations can be solved in terms of the transition densities a_j. Let

$$u(s) = \sum_{j=0}^{\infty} a_j\, s^j \tag{6.5}$$

be the generating function of the sequence $\{a_j\}$.

Theorem 9.11. We have

$$\frac{\partial}{\partial t} F(t, s) = u(s) \frac{\partial}{\partial s} F(t, s) \tag{6.6}$$

and

$$\frac{\partial}{\partial t} F(t, s) = u(F(t, s)). \tag{6.7}$$

Proof: From (6.1) and (6.2) we have

$$
\begin{aligned}
F(h, s) &= \sum_{j=0}^{\infty} p_{1j}(h)\, s^j \\
&= \{1 + a_1 h + o(h)\} s + \sum_{j \neq 1} \{a_j h + o(h)\}\, s^j \\
&= s + h \sum_{j=0}^{\infty} a_j s^j + o(h) = s + hu(s) + o(h).
\end{aligned} \tag{6.8}
$$

Thus, using (6.4a), (6.8) and Taylor's expansion for a function of two variables, we get

$$
\begin{aligned}
F(t+h, s) &= F(t, F(h, s)) \\
&= F(t, s + hu(s) + o(h)) \\
&= F(t, s) + hu(s) \frac{\partial}{\partial s} F(t, s) + o(h).
\end{aligned}
$$

so that

$$\lim_{h \to 0} \frac{F(t+h, s) - F(t, s)}{h} = u(s) \frac{\partial}{\partial s} F(t, s) + \lim_{h \to 0} \frac{o(h)}{h},$$

whence we get (6.6). Again, using (6.4a) and (6.8), we get

$$\begin{aligned} F(h+t, s) &= F(h, F(t, s)) \\ &= F(t, s) + hu(F(t, s)) + o(h). \end{aligned}$$

Thus

$$\lim_{h\to 0} \frac{F(h+t, s) - F(t, s)}{h} = u(F(t, s)) + \lim_{h\to 0} \frac{o(h)}{h}$$

whence we get (6.7). ▲

Each of the equations (6.6) and (6.7) is subject to the initial condition $F(0, s) = s$ (eqn. (6.2a)).

Solution of the equations. $F(t, s)$ can be obtained by solving either the forward equation (6.6) or the backward equation (6.7) with the help of the initial condition $F(0, s) = s$.

Noting that $\frac{d}{dz}\int_0^z f(x)dx = f(z)$, and using $F(0, s) = s$, it can be seen that the equation (6.7) is equivalent to

$$\int_s^{F(t,s)} \frac{dx}{u(x)} = t, \tag{6.9}$$

provided that $u(x) \neq 0$ in $F(0, s) = s \leqslant x \leqslant F(t, s)$. Thus $F(t, s)$ can be obtained theoretically from (6.9).

The forward equation (6.6) together with $F(0, s) = s$ leads to an identical solution.

Example 6 (a). *Simple linear birth and death process.* Here

$$\begin{aligned} \Pr\{X(t+h) = n+k-1 \mid X(t) = n\} &= n\lambda h + o(h), & k &= 2 \\ &= n\mu h + o(h), & k &= 0 \\ &= 1 - n(\lambda+\mu)h + o(h), & k &= 1 \\ &= o(h), & k &\geqslant 3, \end{aligned}$$

i.e. $a_0 = \mu,\ a_1 = -(\lambda+\mu),\ a_2 = \lambda,\ a_k = 0, \quad k \geqslant 3.$

Thus we have

$$u(s) = \lambda s^2 - (\lambda+\mu)s + \mu.$$

From (6.9), we have

$$t = \int_s^{F(t,s)} \frac{dx}{u(x)} = \int_s^{F(t,s)} \frac{dx}{\lambda(x-1)(x-\mu/\lambda)} = \frac{1}{(\mu-\lambda)} \log \frac{(x-\mu/\lambda)}{(x-1)}\bigg|_s^{x=F(t,s)}$$

and on simplification, we get

$$F(t, s) = \frac{\mu(s-1) - (\lambda s - \mu)\, e^{(\mu-\lambda)t}}{\lambda(s-1) - (\lambda s - \mu)\, e^{(\mu-\lambda)t}}. \tag{6.10}$$

Expanding $F(t, s)$ as a power series in s, we can get $p_{1k}(t)$ as coefficient of s^k, $k = 0, 1, 2, \ldots$ [see, Sec. 4.4 under 'Linear growth process'].

Probability of extinction: It may be noted that

$$p_{10}(t) = \Pr\{X(t) = 0 \mid X(0) = 1\} \equiv F(t, 0)$$

is a non-decreasing function of t. The probability of extinction q is defined by

$$q = \lim_{t \to \infty} p_{10}(t). \tag{6.11}$$

Theorem 9.12. The probability of extinction q is the smallest root in [0, 1] of the equation $u(s) = 0$; further, $q = 1$ ($q < 1$) iff $u'(1) \leqslant (>) 0$.

Proof: From (6.7) we find that $p_{10}(t)$ satisfies the equation

$$\frac{d}{dt} p_{10}(t) = u(p_{10}(t)). \tag{6.12}$$

Integrating, we have

$$p_{10}(t) = \int_0^t u(p_{10}(y))\, dy, \tag{6.13}$$

since $\qquad p_{10}(0) = 0.$

We have $u(1) = 0$ and $u(0) \geqslant 0$. If $u(0) = a_0 = 0$, then $p_{10}(0) = 0$ and $q = 0$; then 0 is the smallest non-negative root of $u(s) = 0$, and the theorem holds trivially. Let $u(0) = a_0 > 0$ and q^* be the smallest root of the equation $u(s) = 0$ in $0 \leqslant s \leqslant 1$. In other words, $u(s) = 0$ for $s = q^*$, while $u(s) > 0$ in $0 \leqslant s < q^*$. Now $p_{10}(t)$ is a non-decreasing function of t and so from (6.12) we find $u(p_{10}(t)) \geqslant 0$, for $t > 0$, i.e. $u(p_{10}(t))$ is non-negative for all $t > 0$. Hence $p_{10}(t) \leqslant q^*$ for all $t > 0$. Thus

$$q = \lim_{t \to \infty} p_{10}(t) \leqslant q^*;$$

further $u(q) \geqslant 0$. Suppose, if possible, that $u(q) > 0$, then while $p_{10}(t)$ lies in [0, 1], by the continuity of $u(x)$,

$$\int_0^t u(p_{10}(y))\, dy \to \infty \quad \text{as} \quad t \to \infty;$$

as $\lim_{t \to \infty} p_{10}(t)$ is finite, the above leads to a contradiction of (6.13), which

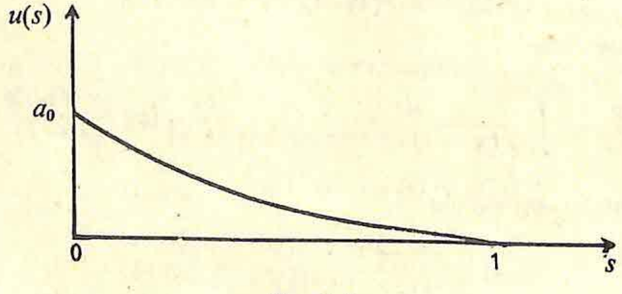


Fig. 9.3

is due to the supposition that $u(q) > 0$. Hence we must have $u(q) = 0$. As $q \leqslant q^*$ and as q^* is the smallest root of $u(s) = 0$, we must have $q = q^*$. We have $0 \leqslant q = q^* \leqslant 1$. Since $u(0) \geqslant 0$. $q = 1$ iff the function $u(s)$ is a decreasing function at $s = 1$, so that $u'(1) \leqslant 0$, if it exists; on the other hand $q < 1$ iff $u'(1) > 0$. Hence the theorem. ▲

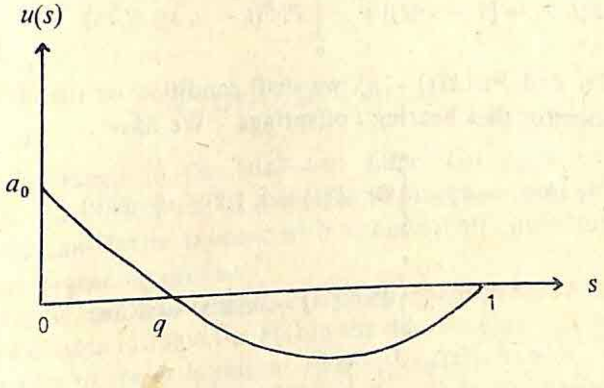


Fig. 9.4

Example 6(a) (contd.). In case of linear birth and death process $u(s) = \lambda\,(s-1)\,(s-\mu/\lambda)$. Thus $q = 1$ iff $\lambda \leqslant \mu$, i.e. iff $u'(1) \leqslant 0$.

9.7 Age Dependent Branching Process: Bellman-Harris process

In the preceding section we considered that the lifetimes of objects (or particles, individuals, organisms) are exponential random variables. Here we shall generalise this further; we shall consider that the lifetimes have general, and not necessarily exponential, distributions.

Suppose that an object (ancestor) at time $t = 0$ initiates the process. At the end of its lifetime it produces a random number of direct descendants having offspring distribution $\{p_k\}$ (with p.g.f. $P(s)$). We assume, as before, that these descendants act independently of each other and that *at the end of its lifetime*, each descendant produces its own descendants with the same offspring distribution $\{p_k\}$, and that the process continues as long as objects are present. Suppose that the lifetimes of objects are i.i.d. random variables with d.f. G (which is also independent of the offspring distribution). Let $\{X(t), t \geqslant 0\}$ be the number of objects alive at time t. The stochastic process $\{X(t), t \geqslant 0\}$ is known as an *age-dependent* (or *general time*) branching process. Such a process is also known as a *Bellman-Harris process*, after Bellman and Harris who first considered such a process in 1948. We shall consider here Bellman-Harris type age-dependent process. An age-dependent process is, in general, not Markovian.

Theorem 9.13 The generating function

$$F(t, s) = \sum_{k=0}^{\infty} \Pr\{X(t) = k\}\, s^k \tag{7.1}$$

of an age-dependent branching process $\{X(t), t \geqslant 0\}$, $X_0 = 1$ satisfies the integral equation

$$F(t, s) = [1 - G(t)]\, s + \int_0^t P(F(t-u, s))\, dG(u). \tag{7.2}$$

Proof: To find $\Pr\{X(t) = k\}$, we shall condition on the lifetime T at which the ancestor dies bearing i offsprings. We have

$$\begin{aligned}\Pr\{X(t) = k\} &= \int_0^{\infty} \Pr\{X(t) = k \mid T = u\}\, dG(u)\\ &= \int_0^{t} \Pr\{X(t) = k \mid T = u\}\, dG(u)\\ &\quad + \int_t^{\infty} \Pr\{X(t) = k \mid T = u\}\, dG(u).\end{aligned}$$

In case of the second term, $u > t$. Given that $T = u$, the number of objects at time t is then still 1 (the ancestor) and the expression under square brackets equals δ_{1k}. Thus the second term yields $\delta_{1k}\{1 - G(t)\}$.

In case of the first term $u \leqslant t$, the ancestor dies at time $u \leqslant t$, leaving $i\,(\in N)$ direct descendants: the probability of this is $p_i dG(u)$; and further these i descendants (who independently initiate processes at time u) leave k objects in the remaining time $t - u$: the probability of this is equal to the coefficient of s^k in the expansion of $[F(t-u, s)]^i$ as a power series in s. Thus we have,

$$\begin{aligned}\Pr\{X(t) = k\} &= \{1 - G(t)\}\, \delta_{1k}\\ &\quad + \int_0^t \sum_{i=0}^{\infty} p_i dG(u)\, \{\text{coeff. of } s^k \text{ in the expansion of } [F(t-u, s)]^i\}.\end{aligned}$$

Multiplying both sides by s^k, $k = 0, 1, 2, \ldots$ and summing over k, we find that the l.h.s. equals $F(t, s)$; and that the first term on the r.h.s equals $\{1 - G(t)\}\, s$ and the second term equals

$$\begin{aligned}&\int_0^t \Big[\sum_{i=0}^{\infty} p_i \Big\{\sum_{k=0}^{\infty} [\text{coeff. of } s^k \text{ in } (F(t-u, s))^i]\, s^k\Big\}\Big]\, dG(u)\\ &= \int_0^t \Big[\sum_{i=0}^{\infty} p_i\, (F(t-u, s))^i\Big]\, dG(u) = \int_0^t P(F(t-u, s))\, dG(u).\end{aligned}$$

Hence the theorem. ▲

Note: The integral equation (7.2) cannot easily be solved in the general case. For such finer questions such as existence and uniqueness of solutions of (7.2) and concerning properties of F, see Harris (1963) or Athreya and Ney (1972).

Particular case: *Exponential lifetime.* Let $G(t) = 1 - e^{-bt}$, then (7.2) yields

$$F(t, s) = se^{-bt} + be^{-bt} \int_0^t P(F(v, s))\, e^{bv}\, dv$$

whence

$$\frac{\partial F}{\partial t} = b\,[P(F(t, s)) - F(t, s)], \tag{7.3}$$

which is equivalent to the backward differential equation (6.7). If $F(t, 1) = 1$, then $\{X(t), t \geqslant 0\}$ becomes a Markov branching process. Thus, a Bellman-Harris process with exponentially distributed lifetime is a Markov branching process.

It can also be seen that when G is a unit step function, (7.2) yields the functional iteration formula for $P(s)$ in the discrete case.

We now turn to the expectation $M(t) = E\{X(t)\}$ and its asymptotic behaviour. $M(t)$ satisfies an integral equation (7.4) which can be obtained rigorously from (7.2). We shall obtain (7.4) directly by using arguments similar to those used to obtain (7.2). Further, we shall show that under certain conditions, $M(t)$ is asymptotically exponential.

Theorem 9.14. The expectation $M(t) = E\{X(t)\}$ of an age-dependent process $\{X(t), t \geqslant 0\}$, $X_0 = 1$ satisfies the integral equation

$$M(t) = [1 - G(t)] + m \int_0^t M(t - u)\, dG(u), \tag{7.4}$$

where $m = P'(1)$ is the mean of the offspring distribution. Further, if $m = 1$, then $M(t) = 1$ is a solution of (7.4). Suppose that α is the unique positive root of

$$\int_0^\infty e^{-\alpha u}\, dG(u) = \frac{1}{m}. \tag{7.5}$$

If $m > 1$ and G is not lattice, then as $t \to \infty$,

$$e^{-\alpha t} M(t) \to \frac{m-1}{\alpha m^2 \int_0^\infty u e^{-\alpha u}\, dG(u)}. \tag{7.6}$$

Proof: By conditioning on the lifetime T of the ancestor, we get

$$M(t) = \int_0^\infty E\{X(t) \mid T = u\}\, dG(u) \tag{7.7}$$

If $u > t$ then $E\{X(t) \mid T = u\} = 1$.

If $u \leqslant t$, then the ancestor lives for time u at the end of which, it leaves i direct descendants with probability p_i; each of these direct descendants initiates a process, the number of objects of such a process having the same distribution as $X(t-u)$. Thus for $u \leqslant t$,

$$E\{X(t) \mid T = u\} = \sum_{i=0}^{\infty} ip_i E\{X(t-u)\}$$

$$= \{M(t-u)\}\, m.$$

Using the above argument we see at once that (7.7) yields (7.4).

Obviously, when $m = 1$, $M(t) = 1$ is a solution of (7.4). Suppose that $m > 1$, and that α exists. Denote

$$G_\alpha(s) = m \int_0^s e^{-\alpha y}\, dG(y);$$

this implies that

$$dG_\alpha(s) = m\, e^{-\alpha s}\, dG(s). \tag{7.8}$$

Multiplying (7.4) by $e^{-\alpha t}$ and writing $v(t) = e^{-\alpha t} M(t)$, we get

$$v(t) = e^{-\alpha t}\,[1 - G(t)] + \int_0^t v(t-u)\, dG_\alpha(u), \tag{7.9}$$

which is a renewal type equation.

Using Theorem 6.3 (Ch. 6) we find that it has the unique solution given by

$$v(t) = e^{-\alpha t}\,\{1 - G(t)\} + \int_0^t e^{-\alpha(t-u)}\,\{1 - G(t-u)\}\, dm_\alpha(u),$$

where $m_\alpha(u)$ is of the form indicated in the theorem.

By using the Key renewal theorem (Theorem 6.7), we get, as $t \to \infty$

$$e^{-\alpha t} M(t) = v(t) \to \frac{\displaystyle\int_0^\infty e^{-\alpha t}\,\{1 - G(t)\}\, dt}{\mu},$$

where

$$\mu = \int_0^\infty u\, dG_\alpha(u) = m \int_0^\infty u e^{-\alpha u}\, dG(u).$$

Integrating by parts the integral in the numerator, we have

$$\int_0^\infty e^{-\alpha t}\,\{1 - G(t)\}\, dt = -\{1 - G(t)\}\frac{e^{-\alpha t}}{\alpha}\Bigg|_{t=0}^{t=\infty} - \left(\frac{1}{\alpha}\right)\int_0^\infty e^{-\alpha t}\, dG(t)$$

$$= \frac{1}{\alpha} - \frac{1}{\alpha m} \quad \text{(using (7.5))}.$$

Thus, as $t \to \infty$

$$e^{-\alpha t}M(t) \longrightarrow \frac{m-1}{\alpha m^2 \int\limits_0^\infty u e^{-\alpha u}\, dG(u)}.$$

Hence the theorem. ▲

Note: A root α of the equation (7.5)

$$\int\limits_0^\infty e^{-\alpha u} dG(u) = \frac{1}{m},$$

when it exists, is known as a *Malthusian parameter*; the constant α represents the geometric rate of increase postulated by Malthus (1766–1834). It is to be noted that when $m \geqslant 1$, there exists a unique positive root α; if $m = 1$, $\alpha = 0$. If $m < 1$, there may or may not be a real root α and if it exists $\alpha < 0$.

Extinction probability is defined by

$$q = \Pr\{Z(t) = 0, \text{ for all sufficiently large } t\}.$$

We state, without proof, the following theorem; for proof and for further results see Athreya and Ney (1972).

Theorem 9.15. If the extinction probability q is 0, then $F(t, 0) = 0$. If $q > 0$, then for each $t \geqslant 0$, $F(t, 0) < q$, and $F(t, 0)$ is a non-decreasing function of t; further $\lim\limits_{t \to \infty} F(t, 0) = q$.

General branching processes: In the Bellman-Harris model it is assumed that an object produces offsprings *at the end of its lifetime*. The offspring distribution is independent of the mother's lifetime.

A model allowing *real* age-dependence was introduced in 1964 by Sevast'yanov, who introduced the possibility of dependence between mother's lifetime and offspring distribution. He considered the generalisation that an object produces offsprings not necessarily at the end of its lifetime but at randomly chosen instants during its lifetime. For results concerning Sevast'yanov type of branching processes, the reader is referred to the books by Jagers (1975) and Sevast'yanov (1971, in Russian; also to his 1964, 1968 papers).

Some Extensions and Generalisations of Continuous Time Processes

In Sec. 9.5, we considered conditioned limit laws of the G.W. processes. Towards the end of the section we indicated some of the extensions of the classical G.W. processes, such as G.W. processes with immigration, G.W. processes in varying and random environments, multitype G.W. processes etc. Such extensions have also been done in the case of continuous time processes. Continuous time multitype processes such as multitype Bellman-Harris process and multitype age-dependent processes have been studied in great details. Continuous time branching processes in varying and random environments, processes with immigration (Pakes, 1971, 1974, 1979), and processes with disasters (Kaplan, 1975) and others have received considerable attention. The study of the conditioned limit laws in various cases has been a topic of great interest. Functional equation technique has led to new results in the theory of Jirina processes, viz. branching processes in discrete time on a continuous state space (Seneta).

Applications of branching processes have been made in several fields like biology, epidemiology, demography and also in nuclear physics and chemistry. The books by Mode, Jagers, Srinivasan, Keyfitz etc. discuss many interesting models of branching processes in such areas, besides covering theoretical developments.

It is not possible to give here references of even the most important works; a limited number of references given here may be of use to the readers.

We have come a long way from the day when Francis Galton (1822–1911) first formulated the basic problem of branching processes in *Educational Times*, London, 1st April 1873, as under:

> *Problem* 4001: 'A large nation, of whom we will concern ourselves with adult males, N in number, and who each bear separate surnames, colonise a district. Their law of population is such that, in each generation, a_0 per cent of the adult males have no male children who reach adult life; a_1 have one such child, a_2 have two, and so on up to a_5, who have five.
>
> Find (1) what proportion of surnames will have become extinct after r generations; and (2) how many instances there will be of the same surname being held by m persons'.

The theory of branching process was formally initiated in the following year by Galton and Watson (1874).

Branching processes have today emerged as a very rich field of research both from the point of view of theoretical structures and of applications which extend to several directions.

EXERCISES

9.1 If $\{X_n, n = 0, 1, 2, \ldots\}$, $X_0 = 1$ is a branching process with offspring p.g.f. $P(s)$, show that $\{X_{rk}, r = 0, 1, \ldots\}$, where k is a fixed positive integer, is also a branching process.

9.2 Suppose that each individual at the end of one unit of time produces either k (where k is a fixed positive integer $\geqslant 2$) or 0 direct descendants with probability p or q respectively. Considering the root of the equation $P(s) = s$, verify that the probability of ultimate extinction is less than one. Find this probability for $k = 3$, when $p = q = \frac{1}{2}$.

9.3 If the offspring distribution in a G.W. process is geometric $\{q^k p\}$, $k = 0, 1, 2, \ldots$, show that

$$P_n(s) = \frac{p(q^n - p^n) - (q^{n-1} - p^{n-1})\, pqs}{(q^{n+1} - p^{n+1}) - (q^n - p^n)\, qs}, \quad p \neq q$$
$$= \frac{n - (n-1)\, s}{(n+1) - ns}, \quad p = q.$$

9.4 Find the variance of X_n (i.e. the number of objects in nth generation) in a G.W. process from the relation $P_n = P(P_{n-1}(s))$.

9.5 Show that the m.g.f. $\varphi_n(s) = E\{e^{-sW_n}\}$, $n = 1, 2, \ldots$, satisfies the relations

$$\varphi_{n+1}(ms) = P\,[\varphi_n(s)].$$

9.6 Show that the p.g.f. of the conditional distribution of X_n given $X_n > 0$ is

$$\frac{P_n(s) - P_n(0)}{1 - P_n(0)}.$$

Find $\Pr\{X_n = r \mid X_n > 0\}$ when Pr {no. of offspring $= k\} = (\frac{1}{2})^{k+1}$, $k = 0, 1, 2, \ldots$.

9.7 For a Galton-Watson process, show that the expected number of total progeny is given by

$$E(Y_n) = (1 - m^{n+1})/(1 - m), \text{ if } m \neq 1$$
$$= n + 1, \text{ if } m = 1$$

and that, if $m < 1$, $E(Y_n) \to \dfrac{1}{1-m}$ as $n \to \infty$.

Show that

$$\text{var}\,(Y_n) = (2n + 1)(n + 1)\, n\sigma^2/6, \text{ if } m = 1.$$

9.8 Show that for a binary splitting process (i.e. with offspring distribution $p_0 = q$, $p_2 = p = 1 - q$), if $p < \frac{1}{2}$

$$H(s) = \frac{1 - \sqrt{4s^2pq}}{2sp}$$

(where $H(s) = \lim\limits_{n\to\infty} R_n(s)$, $R_n(s) = E\{s^{Y_n}\}$, Y_n = total progeny up to nth generation). Deduce that

$$\Pr\{Y_\infty = 2k\} = 0$$
$$\Pr\{Y_\infty = 2k - 1\} = \frac{(-1)^{k-1}}{2p} \binom{\frac{1}{2}}{k} (4pq)^k.$$

9.9 Find the limiting distribution of the total progeny in a G.W. process with geometric offspring distribution $\{q^k p\}$, $k = 0, 1, 2, \ldots$.

9.10 Consider a G.W. process with geometric offspring distribution $\{q^k p\}$, $k = 0, 1, 2, \ldots$. Verify that when $m = 1$

$$\lim_{n\to\infty} n \Pr\{X_n > 0\} = \frac{2}{\sigma^2}.$$

When $m < 1$ $(q < p)$, verify that Yaglom's theorem holds. Find

$$b_k = \lim_{n\to\infty} \Pr\{X_n = k \mid X_n > 0\}, \quad k = 1, 2, \ldots$$

and verify that $B(s) = \sum_{k=1}^{\infty} b_k s^k$ satisfies the equation

$$B(P(s)) = mB(s) + 1 - m.$$

9.11 For a G.W. process with $m < 1$, $\sigma^2 < \infty$, if $B(s)$ is the the p.g.f. of the distribution $\{b_k\}$ defined by by (5 8), show that $B''(1) < \infty$. Further, show that the variance $\sigma_b{}^2$ of $\{b_k)$ is given by

$$\sigma_b{}^2 = \frac{\zeta\sigma^2}{m - m^2} - \zeta^2$$

where $$\zeta = \sum_{k=1}^{\infty} kb_k = \frac{1}{\varphi(0)}.$$

9.12 *Time to extinction* in a G.W. process. Show that the distribution of time to extinction T (which is the first passage time to state 0) is given by

$$r_n = \Pr\{T = n\} = P_n(0) - P_{n-1}(0), \; n = 1, 2, \ldots$$

Find r_n when

(i) $p_k = q^k p, \; k = 0, 1, 2, \ldots$

(ii) $p_k = (1 - a)(1 - c)\, c^{k-1}, \; k = 1, 2, \ldots$

$\quad = a \qquad\qquad , \; k = 0.$

9.13 Suppose that at the end of exponential lifetime (having parameter a) each particle splits into $k(\geqslant 2)$ particles. Show that the generating function $F(t, s)$ of $\{X(t) \mid X_0 = 1\}$ is given by

$$F(t, s) = se^{-at}\,[1 - \{1 - e^{-a(k-1)t}\}\, s^{k-1}]^{-1/(k-1)}.$$

9.14 Find $\text{var}(X(t))$, where $X(t)$ is a Markov branching process with $X(0) = 1$ and having $u(s)$ as the generating function of the transition densities.

9.15 Consider an age-dependent process with exponential lifetime given by $G(t) = 1 - e^{-bt}$. Suppose that at the time of its death an object splits into two similar objects. The process then reduces to the pure birth process considered by Yule. Obtain the differential equation for the generating function of the process. Solving the equation, show that

$$\Pr\{X(t) = k\} = e^{-bt}\,(1 - e^{-bt})^{k-1}, \; k = 1, 2, \ldots.$$

9.16 Starting with an age-dependent process show that the differential equation satisfied by a simple linear birth and death process reduces to

$$\frac{\partial F}{\partial t} = \lambda F^2 - (\lambda + \mu) F + \mu,$$

where λ, μ are the birth and death rates respectively.

REFERENCES

K.B. Athreya, "Some results on multitype continuous time Markov branching processes," *Ann. Math. Stat.*, **39**, 347–357 (1968).

K.B. Athreya and S. Karlin, "On branching processes in random environments; I : Extinction probabilities, and II : Limit theorems, *Ann. Math. Stat.*, **42**, 1499–1520 and 1843–1858 (1971).

———, "Branching processes with random environments", *Proc. Am. Math. Soc.*, **76**, 865–870 (1970).

K.B. Athreya, and P.E. Ney, *Branching Processes*, Springer-Verlag, New York (1972).

R.E. Bellman and T.E. Harris, "On the theory of age dependent stochastic branching processes," *Proc. Nat. Ac. Sc.*, **34**, 601–604 (1948).

———, "On age-dependent binary branching processes," *Ann. Math.*, **55** 280–295 (1952).

D.H. Fearn, "Galton-Watson processes with generation dependence," *Proc. Sixth Berkeley Symp. Math. Stat. and Prob.*, **IV**, 159–172 (1972).

F. Galton and H.W. Watson "On the probability of extinction of families." *J. of Anthr. Inst.*, **6**, 138–144 (1874).

T.E. Harris, *The Theory of Branching Processes*, Springer-Verlag, Berlin (1963).

C.R. Heathcote, "A branching process allowing immigration," *J. Roy. Stat. Soc.* **B.**, **27**, 138–143 (1965); Corrections and comments, **B. 28** 213–217 (1966).

P. Jagers, Galton-Watson processes in varying environments, *J. Appl. Prob.*, **11**, 174–178 (1974).

———, *Branching Processes with Biological Applications*, Wiley, New York, (1975).

A. Joffe, "On the Galton-Watson processes with mean less than one," *Ann. Math. Stat.*, **38**, 264–266 (1967).

S. Karlin, *A First Course in Stochastic Processes*, Academic Press, New York, (1966).

N. Kaplan, A. Sudbury and T. Nilsen, "A branching process with disasters," *J. Appl. Prob.*, **12**, 47–59 (1975).

N. Keyfitz, *Introduction to the Mathematics of Population*, Addison-Wesley, Reading, Mass, (1968).

C.J. Mode, *Multitype Branching Processes*, Elsevier, New York (1971).

A.G. Pakes, "Branching processes with immigration," *J. Appl. Prob.*, **8**, 32–42 (1971).

———, "On the critical G.W. processes with immigration," *J. Austr. Math. Soc.*, **13**, 277–290 (1972).

———, "On Markov branching processes with immigration," *Sankhyā*, A **37**, 129–138 (1974).

———, "Limit theorems for the simple branching processes allowing immigration: I, the case of finite offspring mean," *Adv. Appl. Prob.* **11**, 31–62 (1979).

A.G. Pakes and N.L. Kaplan, "On the subcritical Bellman-Harris process with immigration," *J. Appl. Prob.*, **11**, 652–668 (1974).

E. Seneta, "Functional equations and the Galton-Watson process," *Adv. Appl. Prob.*, **1**, 1–42 (1969).

———, "An explicit limit theorem for the critical Galton-Watson process with immigration," *J. Roy. Stat. Soc.*, ***B***, **32**, 149–152 (1970).

B.A. Sevast'yanov, "Age-dependent branching processes," *Theo. Prob. Appl.*, **9**, 521–537 (1964).

———, "Limit theorems for age-dependent branching processes," *Theo. Prob. Appl.*, **13**, 237–259 (1968).

———, *Branching Processes* (in Russian), Nauka, Moscow (1971).

W.L. Smith, "Necessary conditions for almost sure extinction of a branching process with random environment," *Ann. Math. Stat.*, **39**, 2136–2140 (1968).

W.L. Smith and W.E. Wilkinson, "On Branching processes in random environments," *Ann. Math. Stat.*, **40**, 814–827 (1969).

S.K. Srinivasan, *Stochastic Theory and Cascade Processes*, Elsevier, New York, (1969).

S.K. Srinivasan and K.M. Mehata, *Stochastic Processes*, Tata McGraw-Hill, New Delhi, (1976).

W.C. Torrez, "The birth and death chain in a random environment: instability and extinction theorems," *Ann. Prob.*, **6**, 1026–1043 (1978).

W.E. Wilkinson, "On calculating extinction probabilities for branching processes in random environments," *J. Appl. Prob.* **6**, 478–492 (1969).

Y.S. Yang, "On branching processes allowing immigration," *J. Appl. Prob.*, **9**, 24–31 (1972).

CHAPTER 10

STOCHASTIC PROCESSES IN QUEUEING AND RELIABILITY

10.1 Queueing Systems: General Concepts

The queueing theory had its origin in 1909, when A.K. Erlang (1878–1929) published his fundamental paper relating to the study of congestion in telephone traffic. Erlang is appropriately called the father of queueing theory (for his life and work, see Brockmeyer *et al.* (1948)). The literature on the theory of queues and on the diverse areas of its applications have grown tremendously over the years. The standard works and some of the other publications specifically referred to in our discussions are listed in the references. For complete references, the standard works may be consulted. We consider here only the fundamental aspects of the theory.

A queue or waiting line (*'file d'attente'* in French and *'Warteschlange'* in German) is formed when units (or customers, clients) needing some kind of service arrive at a service channel (or counter) which offers such facility. A queueing system can be described by the flow of units for service, forming or joining the queue, if service is not immediately available, and leaving the system after being served (or sometimes without being served). The basic features which characterise a system are; (i) the input, (ii) the service mechanism, (iii) the queue discipline and (iv) the number of service channels.

By units, we mean those demanding service, e.g. customers at a bank counter or at a reservation counter, calls arriving at a telephone exchange, vehicular traffic at a traffic intersection, machines for repair before a repairman, airplanes waiting for take-off at a busy airport, merchandise waiting for shipment at a yard, computer programmes waiting to be run on a time-sharing basis etc.

The *input* describes the manner in which units arrive and join the system. The system is called a *delay* or *loss system* depending on whether a unit who, on arrival, finds the service facility occupied, joins or leaves the system. The system may have either a limited or an unlimited capacity for holding units. The source from which the units come may be finite or infinite. A unit may arrive either singly or in a group. The interval between two consecutive arrivals (or between two consecutive arrival groups, in case of bulk arrivals) is called the *interarrival time* or *interval.*

The service mechanism describes the manner in which service is rendered. A unit may be served either singly or in a batch. The time required for servicing a unit (or a group, in case of batch service) is called the *service time*.

The queue discipline indicates the way in which the units form a queue and are served. The usual discipline is first come first served (FCFS) or first in first out (FIFO), though sometimes, other rules, such as, last come first served or random ordering before service are also adopted.

The system may have a *single channel* or *s-parallel channels* for service.

The interarrival and service times may be deterministic or chance-dependent. The case when both the interarrival and service times are deterministic is trivial. We shall be generally concerned with chance-dependent interarrival and service times, and the theory will be essentially stochastic. When chance-dependent, the interarrival times between two consecutive arrivals are assumed to be i.i.d. random variables; the service times of units are also assumed to be i.i.d. random variables. Further the two distributions are also taken to be independent.

The mean arrival rate, usually denoted by λ, is the mean number of arrivals per unit time. Its reciprocal is the mean of the interarrival time distribution. The mean service rate, usually denoted by μ, is the mean number of units served per unit time, its reciprocal being the mean service time. In a single channel system, the ratio

$$\rho = \frac{\lambda}{\mu}$$

is called the *traffic intensity*. Though dimensionless, it is expressed in *erlangs* (in honour of A.K. Erlang). The definition of traffic intensity is suitably modified in case of s-channel systems and in case of bulk or batch queueing systems.

Queueing Processes

The following random variables or families of random variables which arise in the study provide important measures of performance and effectiveness of a stochastic queueing system.

(1) The number $N(t)$ in the system at time t, i.e. the number at time t waiting in the queue including the one being served, if any.

(2) The busy period which means the duration of the interval from the moment the service commences with arrival of an unit at an empty counter to the moment the server becomes free for the first time.

(3) The waiting time in the queue (system), i.e. the duration of time a unit has to spend in the queue (system); also the waiting time W_n of the nth arrival.

(4) The virtual waiting time $W(t)$, i.e. the interval of time a unit would have to wait in the queue, were it to arrive at the instant t.

One needs to have their complete probabilistic description. It is clear that $\{N(t), t \geqslant 0\}$, $\{W(t), t \geqslant 0\}$, $\{W_n, n \geqslant 0\}$ are stochastic processes, the first two being in continuous time and the third one in discrete time. It will be seen that some of the queueing processes that we would come across are Markovian and some are semi-Markovian. From some of the non-Markovian processes that arise, Markov chains can be extracted at suitable regeneration points and semi-Markov processes can be constructed there from. The theory of Markov chains and semi-Markov processes thus plays an important role in the study of queueing processes.

Notation: A very convenient notation designed by Kendall (1951) to denote queueing systems has been universally accepted and used. It consists of a three-part descriptor $A/B/C$, where the first and the second symbols denote the interarrival and service time distributions respectively, and the third denotes the number of channels or servers. A and B usually take one of the following symbols:

M : for exponential (Markovian) distribution
E_k : for Erlang-k distribution
G : for arbitrary (General) distribution
D : for fixed (Deterministic) interval

Thus, by $M/G/1$ is meant a single channel queueing system having exponential interarrival time distribution and arbitrary (general) service time distribution. By $M/G/1/R$ is meant the same system with the fourth descriptor R denoting that the system has a limited holding capacity R.

Steady State Distribution

$N(t)$, the number in the system at time t and its probability distribution, given by

$$p_n(t) = \Pr\{N(t) = n \mid N(0) = \cdot\}$$

are both time dependent. For a complete description of the queueing process we need consider transient or time-dependent solutions. It is often difficult to obtain such solutions. Further, in many practical situations, one needs to know the behaviour in steady state, i.e. when the system reaches an equilibrium state, after being in operation for a pretty long time.

It is easier and convenient to determine

$$p_n = \lim p_n(t) \text{ as } t \to \infty.$$

provided the limit exists. It is necessary to know the condition for the existence of the limit in the first place. This will be discussed in due course. When the limit exists, it is said that the system has reached *equilibrium* or *steady state* and the problem then boils down to finding the steady state solutions.

Some General Relationships in Queueing Theory

There are certain useful statements and relationships in queueing theory which hold under fairly general conditions. Though rigorous mathematical proofs of such relations are somewhat complicated, intuitive and heuristic proofs are simple enough and have been known for long (Morse (1947)). It has been argued (Krakowski, (1973) also Stidham (1974) that conservation methods could very well be applied to supply proofs of some of these relations. Conservation principles have played a fundamental role in physical and engineering sciences as well as in economics etc. Similar principles may perhaps be applied in obtaining relations for queueing systems in steady state. Some such relations are given below. The most important one is

$$L = \lambda W \tag{1.1}$$

where λ is the arrival rate, L is the expected number of units in the system and W is the expected waiting time in the system in steady state. A rigorous proof of the relation has been given by Little (1961) and so the relation is known as *Little's* formula. This result, of great generality, is independent of the form of interarrival and service time distributions, and holds under some very general conditions.

Denote the expected number in the queue and the expected waiting time in the queue in steady state by L_Q and W_Q respectively. These are related by a similar formula:

$$L_Q = \lambda W_Q. \tag{1.2}$$

(see also Brumelle for a generalisation of (1.1)).

A relation which holds for $G/G/1$ queue in steady state is

$$\lambda = (1 - p_0)\mu \quad \text{or} \quad p_0 = 1 - \rho. \tag{1.3}$$

There are other relations, which hold under somewhat restrictive conditions. For example, for a system with Poisson input, the Pollaczek-Khinchine formula holds. We shall prove this formula later, while proof of Little's formula will be omitted.

10.2 The Queueing Model M/M/1: steady state behaviour

This single server model envisages Poisson input and exponential service time with FCFS queue discipline. Such a queue is also known as the *simple queue* (or *Poisson queue* or *Markovian queue*).

The arrivals occur in accordance with a Poisson process with intensity λ (say), i.e. the probability that an arrival occurs in an infinitesimal interval of length h is $\lambda h + o(h)$, while that of more than one arrival is $o(h)$. This is equivalent to the statement that the distribution of the interarrival times is exponential having p.d.f. $a(t) = \lambda e^{-\lambda t}$. The distribution of the service times is exponential, with parameter μ (say) having p.d.f. $b(t) = \mu e^{-\mu t}$. In other words, the probability that one service is completed in an interval of infinitesimal length h is $\mu h + o(h)$, while the probability of more than one completion of service is $o(h)$. It is assumed that interarrival times as well as service times are stochastically independent. We shall call such a queue an $M/M/1$ queue with parameters λ, μ

(or a simple queue with parameters λ, μ). The mean interarrival time is $1/\lambda$ and the mean service time is $1/\mu$. The ratio $\rho = \lambda/\mu$ is the *traffic intensity* (also called *load factor* or *utilization factor*).

Let $N(t)$ be the number in the system (those in the queue including the one being served, if any) at instant $t > 0$; then $\{N(t), t \geqslant 0\}$ is a Markov process in continuous time with denumerable number of states $\{0, 1, 2, \ldots\}$. Here transitions take place only to two neighbouring states. This is a type of birth and death process.

The arrivals here can be thought of as "births" and the service completions as "deaths". The birth and death process with $\lambda_n = \lambda, n \geqslant 0$ and $\mu_n = \mu, n \geqslant 1, \mu_0 = 0$, is also known as *immigration-emigration* process (see Sec. 4.4). Let $\Pr\{N(t) \quad n \mid N(0) = \cdot\} = p_n(t), n \geqslant 0$. Using the same probabilistic arguments (or directly from equations (4.6) and (4.4) of Sec. 4.4) we at once get

$$p_0'(t) = -\lambda p_0(t) + \mu p_1(t) \tag{2.1}$$

$$p_n'(t) = -(\lambda + \mu)\, p_n(t) + \lambda p_{n-1}(t) + \mu p_{n+1}(t), \; n \geqslant 1. \tag{2.2}$$

A. Steady-State Solution

First, let us assume the existence of a "steady-state". Then as t tends to infinity, $p_n(t)$ tends to a limit p_n, independent of t. The equations of steady-state probabilities p_n can be obtained by putting $p_n'(t) = 0$ and and replacing $p_n(t)$ by p_n in (2.1) and (2.2). We get

$$0 = -\lambda p_0 + \mu p_1 \tag{2.3}$$

$$0 = -(\lambda + \mu)\, p_n + \lambda p_{n-1} + \mu p_{n+1}, \; n = 1, 2, \ldots. \tag{2.4}$$

Dividing by μ, and writing ρ for λ/μ, the above equations can be written as

$$p_1 - \rho p_0 = 0 \tag{2.3a}$$

$$p_{n+1} - (1 + \rho) p_n + \rho p_{n-1} = 0, \; n = 1, 2, \ldots. \tag{2.4a}$$

The solution of these equations was considered in Example 5(b) Ch. 1. The solution is given by

$$p_n = p_0 \rho^n, \; n \geqslant 0. \tag{2.5}$$

Now in order that $\{p_n\}$ is a probability distribution, we must have

$$\sum_{n=0}^{\infty} p_n = 1 \text{ or, } p_0 \sum_{n=0}^{\infty} \rho^n = 1.$$

The infinite geometric series converges to the sum $1/(1 - \rho)$ *iff* $\rho < 1$. For existence of steady state solution, ρ *must be less than* 1. We have then

$$p_n = (1 - \rho)\, \rho^n, \; n \geqslant 0. \tag{2.6}$$

The distribution $\{p_n\}$ of the random variable N, the number of the system in steady state, is geometric (see also Example 4(c) Ch. 7, where the distribution was obtained by using the theory of semi-Markov processes).

The probability that the system is empty is $p_0 = 1 - \rho$ and that the system is not empty is given by $\Pr\{N \geqslant 1\} = 1 - p_0 = 1 - (1 - \rho) = \rho$. We have

$$E\{N\} = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}$$

and
$$\operatorname{var}(N) = \frac{\rho}{(1-\rho)^2} = \frac{\lambda\mu}{(\mu-\lambda)^2}.$$

The following table gives values of $E\{N\}$, var (N) for some values of ρ:

ρ	0.1	0.3	0.5	0.7	0.8	0.9	0.95	0.99
$E\{N\}$	0.11	0 43	1.00	2.33	4.00	9.00	19.00	99.00
var(N)	0.12	0.61	2.00	7.78	20.00	90.00	380.00	9900.00

Note: $\rho < 1$ (or equivalently $\lambda < \mu$) is a necessary condition for the existence of steady state, and that the queue discipline has no effect on the distribution of N. The queue discipline will have to be taken account of in obtaining the following distributions.

Waiting Time Distributions

Waiting Time in the Queue

We assume that the queue discipline is FCFS. Let the random variable W_Q denote the time spent waiting in the queue (or '*queueing time*') of a test unit. Now $W_Q = 0$ when there is no unit in the system, the probability for which is $p_0 = 1 - \rho$. Given that there are n $(\geqslant 1)$ units in the system, $W_Q = S_n$, where

$$S_n = v_1' + v_2 + \ldots + v_n,$$

v'_1 being the residual service time of the customer being served and $v_2, v_3, \ldots, v_n$ being the service times of the units in the queue. Now v'_1, the residual of an exponential distribution with mean $1/\mu$ is again an exponential distribution with the same mean. Thus S_n is the sum of n i.i.d. exponential distributions and is a variable with gamma distribution having p.d.f.

$$\frac{\mu^n x^{n-1} e^{-\mu x}}{\Gamma(n)}, \; x > 0$$

(see Theorem 1.10).

Hence we have, for $x > 0$,

$$\begin{aligned} w_q(x)\,dx &= \Pr\{x \leqslant W_Q < x + dx\} \\ &= \sum_{n=1}^{\infty} \Pr\{x \leqslant S_n < x + dx \mid \text{the test unit finds } n \text{ in the system}\} \\ &\qquad \cdot \Pr\{\text{the test unit finds } n \text{ in the system}\} \\ &= \sum_{n=1}^{\infty} \left\{\frac{\mu^n x^{n-1} e^{-\mu x}}{\Gamma(n)}\,dx\right\} p_n \end{aligned}$$

$$= \mu e^{-\mu x}(1-\rho)\,\rho\left\{\sum_{n=1}^{\infty}\frac{(\mu\rho x)^{n-1}}{(n-1)!}\right\}dx$$

$$= \mu\rho\,(1-\rho)\,e^{-\mu(1-\rho)x}\,dx,\ x>0.$$

Thus the p.d.f. of the distribution is given by

$$\begin{aligned} w_q(x) &= 0, \qquad x=0 \\ &= \mu\rho\,(1-\rho)\,e^{-\mu(1-\rho)x},\ x>0. \end{aligned} \tag{2.7}$$

The distribution function $F(x)$ of W_Q is given by

$$F(x) = 1-\rho,\ x=0$$

and

$$\begin{aligned} F(x) &= \int_0^x w_q(t)\,dt \\ &= 1-\rho\,e^{-\mu(1-\rho)x},\ x>0. \end{aligned} \tag{2.8}$$

The above gives the (unconditional) distribution of the waiting time of a test unit. The probability that the unit has not to wait is $1-\rho$ and that it has to wait is ρ. We have

$$\begin{aligned} E(W_Q) &= 0\cdot(1-\rho)+\int_{0+}^{\infty}\mu\rho\,(1-\rho)x\,e^{-\mu(1-\rho)x}\,dx \\ &= \frac{\rho}{\mu\,(1-\rho)} = \frac{\lambda}{\mu\,(\mu-\lambda)}. \end{aligned} \tag{2.9}$$

Now the (conditional) distribution of the waiting time in the queue denoted by ${}_cW_Q$ given that the test unit has to wait, is given by

$$\begin{aligned} f_c(x)\,dx &= \Pr\{x \leqslant {}_cW_Q < x+dx\},\ x>0 \\ &= \Pr\{x \leqslant W_Q < x+dx \mid \text{the test unit has to wait}\} \\ &= \frac{\mu\rho\,(1-\rho)\,e^{-\mu(1-\rho)x}}{\rho}\,dx = \mu(1-\rho)\,e^{-\mu(1-\rho)x}dx,\ x>0. \end{aligned}$$

The distribution of ${}_cW_Q$ is exponential (with mean $1/\mu\,(1-\rho)$) while that of W_Q is modified exponential.

Waiting Time in the System

The random variable 'time spent in the system' (or *sojourn time*) by a unit includes the service time of the unit besides its queueing time. This variable is given $W_S = W_Q + S$, where W_Q is the queueing time and S is the service time, and its distribution can be obtained by convolution. We give below a derivation similar to the one given for W_Q.

Now the test unit has to wait in the system even if the system is empty ($n=0$), the time spent being equal to his service time. Given that there are $n(\geqslant 0)$ units in the system, its waiting time is $W_S = S_{n+1}$, where $S_{n+1} = v'_1 + v_2 + \ldots + v_{n+1}$, v_{n+1} being its own service time.

Thus the p.d.f. $w_s(x)$ of W_S for $x > 0$ is given by

$$\begin{aligned} w_s(x)\,dx &= \Pr\{x \leqslant W_S < x + dx\} \\ &= \sum_{n=0}^{\infty} \Pr\{x \leqslant S_{n+1} < x + dx \mid \text{given that there are } n \text{ units in the system}\}\cdot \Pr\{N = n\}. \\ &= \left\{ \sum_{n=0}^{\infty} \mu \frac{(\mu x)^n}{\Gamma(n+1)} e^{-\mu x}\,dx \right\} \{(1-\rho)\,\rho^n\} \\ &= \mu\,(1-\rho)\,e^{-\mu(1-\rho)x}\,dx. \end{aligned} \tag{2.10}$$

Thus W_S has an exponential distribution with mean $1/\{\mu(1-\rho)\} = 1/(\mu - \lambda)$ and variance $1/\{\mu\,(1-\rho)\}^2$.

The distribution function is

$$F(x) = \Pr\{W_S \leqslant x\} = \int_0^x w(x)dx = 1 - e^{-\mu(1-\rho)x},\ x > 0.$$

Note: (1) S_{n+1} is the random (geometric) sum of $(n+1)$ i.i.d. exponential variables. That its distribution will be again exponential was also shown in Example 3(g) Ch. 1.

(2) It can be easily verified that Little's formula holds.

Example 2(a). The arrivals at a counter in a bank occur in accordance with a Poisson process at an average rate of 8 per hour. The duration of service of a customer has exponential distribution with a mean of 6 minutes. Find the probability that an arriving customer: (i) has to wait on arrival, (ii) finds 4 customers in the queue, and (iii) has to spend less than 15 minutes in the bank. Estimate also the fraction of the total time that the counter is busy.

Here $\lambda = 8/\text{hr.},\quad \mu = 10/\text{hr.},\quad \rho = 0.8.$

The probability that the customer (i) has to wait is given by $\Pr\{N \geqslant 1\} = \rho = 0.8$, and (ii) finds 4 customers is given by $p_4 = (1-\rho)\,\rho^4 = (0.2)\,(0.8)^4 = 0.08192$.

The probability that he has to wait less than 15 minutes ($= 0.25$ hr.) is given by

$$\begin{aligned} F(0.25) &= 1 - \exp\{-\mu\,(1-\rho)\,(0.25)\} && \text{(from 2.10)} \\ &= 1 - \exp\{-10\,(0.2)\,(0.25)\} = 1 - e^{-0.5} = 0.3935. \end{aligned}$$

The utilization factor $\rho = 0.8$ gives an estimate of the fraction of the total time that the counter is busy.

B. Queue with Limited Waiting Space: The queueing model M/M/1/R

In the $M/M/1$ model, it is assumed that the system can accommodate any number of units. In practice this may seldom be the case. We have thus to consider the situation such that the system has limited waiting space and can hold a maximum number of R units (including the one being served). The analysis is similar. Equation (2.1) will hold but (2.2) will hold only for $1 \leqslant n \leqslant (R-1)$. Further, there will be another relation. Suppose that the number of units in the system in an interval $(0, t)$ is R, then the probability that there will be R units in the interval $(0, t+h)$ is given by

$$p_R(t+h) = p_R(t)\,(1-\mu h) + p_{R\text{-}1}\,(\lambda h)\,(1-\mu h) + o(h)$$

which gives, as $h \to 0$

$$p_R'(t) = -\mu\, p_R(t) + \lambda\, p_{R-1}(t).$$

Thus the steady state equations are

$$\begin{aligned} 0 &= -\lambda\, p_0 + \mu\, p_1 \\ 0 &= -(\lambda+\mu)\, p_n + \lambda\, p_{n-1} + \mu\, p_{n+1},\ n = 1, 2, \ldots, R-1 \\ 0 &= -\mu\, p_R + \lambda\, p_{R-1}. \end{aligned} \tag{2.11}$$

From the first two equations we get, as in (2.5)

$$p_n = p_0\, \rho^n,\ 0 \leqslant n \leqslant R-1.$$

From the last one, $p_R = \rho p_{R-1} = p_0 \rho^R$, so that

$$p_n = p_0 \rho^n \text{ for } n = 0, 1, \ldots, R.$$

Using $\sum_{n=0}^{R} p_n = 1$, we get

$$\begin{aligned} p_0 &= \frac{1-\rho}{1-\rho^{R+1}},\ \rho \neq 1 \\ &= \frac{1}{R+1}, \quad \rho = 1. \end{aligned}$$

Finally, we have, for $n = 0, 1, \ldots, R$,

$$\begin{aligned} p_n &= \frac{(1-\rho)\,\rho^n}{1-\rho^{R+1}},\ \rho \neq 1 \\ &= \frac{1}{R+1}, \quad \rho = 1. \end{aligned} \tag{2.12}$$

The distribution is truncated geometric for $\rho \neq 1$ and uniform for $\rho = 1$. The mean number of units in the system L_R is given by

$$L_R = \sum_{n=0}^{R} n\, p_n.$$

For $\rho = 1$,
$$L_R = \frac{R(R+1)/2}{R+1} = \frac{R}{2}.$$

For $\rho \neq 1$,
$$\begin{aligned} L_R &= \frac{(1-\rho)\,\rho}{1-\rho^{R+1}} \sum_{n=0}^{R} n\, \rho^{n-1} \\ &= \frac{(1-\rho)\,\rho}{1-\rho^{R+1}} \sum_{n=0}^{R} \frac{d}{d\rho}\,(\rho^n) \end{aligned}$$

$$= \frac{(1-\rho)\rho}{1-\rho^{R+1}} \cdot \frac{d}{d\rho}\left\{ \sum_{n=0}^{R} \rho^n \right\}$$

$$= \frac{(1-\rho)\rho}{1-\rho^{R+1}} \cdot \frac{1-(R+1)\rho^R + R\rho^{R+1}}{(1-\rho)^2} \tag{2.13}$$

It is to be noted that since the finite series $\sum_{n=0}^{R} \rho^n$ has a sum for all values of ρ ($\rho < 1$ and $\rho \geqslant 1$), the steady state solution exists even for $\rho \geqslant 1$. Some of the results for $M/M/1$ model (or $M/M/1/\infty$ model) can be obtained by taking limits of the corresponding expressions as $R \to \infty$. But then this could be done only when $\rho < 1$.

10.3 Transient Behaviour of M/M/1 Model

We have now to solve the equations (2.1) and (2.2) in transient state. A number of methods have been put forward for their solutions: the spectral method of Lederman and Reuter (1956); the method of generating function of Bailey (1956); the combinatorial method of Champernowne (1956). The methods are given in all advanced level books on Queueing theory (e.g. Prabhu, Saaty, etc.). We consider Conolly's (1958) difference equation technique and also outline Bailey's (1956) generating function method.

Difference Equation Technique

For obtaining the solution in steady state the number in the state at the initial stage is of no consequence. Here the knowledge of the initial state is needed.

We assume, for simplicity, that the number of units at the instant $t = 0$ is 0. Let

$$p_n(t) = \Pr\{N(t) = n \mid N(0) = 0\}, \quad p_0(0) = 1,\ p_n(0) = 0,\ n \neq 0.$$

We recall the approach to solution of the equations (2.1) and (2.2) as considered in Example 6(c) Ch. 1.

Let $f_n(s)$ be the L.T. of $p_n(t)$. Then from (2.1) and (2.2), we get

$$(s+\lambda) f_0(s) = 1 + \mu f_1(s) \tag{3.1}$$

and

$$(s+\lambda+\mu) f_n(s) = \lambda f_{n-1}(s) + \mu f_{n+1}(s), \quad n \geqslant 1. \tag{3.2}$$

The solution of the equation (3.2) is

$$f_n(s) = A\alpha^n + B\beta^n, \quad n \geqslant 1, \tag{3.3}$$

where A, B are functions of s (and not of n) and α, β are the roots of the characteristic equation

$$\mu x^2 - (s+\lambda+\mu)x + \lambda = 0. \tag{3.4}$$

We are now to find the values of A and B.

Writing $K(s) \equiv K = \{(s + \lambda + \mu)^2 - 4\lambda\mu\}^{1/2}$, we get

$$\begin{matrix}\alpha \\ \beta\end{matrix} \equiv \frac{(s + \lambda + \mu) \pm K}{2\mu} \quad (\alpha \equiv \alpha(s), \beta \equiv \beta(s)). \tag{3.4a}$$

It can be seen that, for all $|s|$ and real and positive λ, μ,

$$|\alpha| > 1 \quad \text{and} \quad |\beta| < 1 \quad \text{(see note below).}$$

Now $\quad \sum\limits_{n=0}^{\infty} p_n(t) = 1$ and hence $\sum\limits_{n=0}^{\infty} f_n(s) = \dfrac{1}{s}$.

Since $|\alpha| > 1$, the convergence of $\sum\limits_{n=0}^{\infty} f_n(s) = \sum\limits_{n=0}^{\infty} (A\alpha^n + A\beta^n)$

implies that $A = 0$. Thus $f_n(s) = B\beta^n$, $n \geqslant 1$. Choose B such that this holds for $n = 0$ also.

We get $\quad f_0(s) = B$, so that

$$f_n(s) = f_0(s)\, \beta^n, \; n \geqslant 0. \tag{3.5}$$

Again, $\quad f_0(s) = \dfrac{1 + \mu f_1(s)}{s + \lambda} = \dfrac{1 + \mu f_0(s)\beta}{s + \lambda}$ (using (3.1) & (3.5))

whence, using $\quad \mu\beta^2 - (s + \lambda + \mu)\beta + \lambda = 0$, we get

$$f_0(s) = \frac{1}{s + \lambda - \mu\beta} = \frac{1 - \beta}{s}.$$

Thus

$$f_n(s) = f_0(s)\, \beta^n = \frac{(1 - \beta)\, \beta^n}{s}. \tag{3.5a}$$

$p_n(t)$ can be obtained by taking the inverse Laplace transform of $f_n(s)$. This is done by using some properties of Laplace transforms and by taking recourse to ingenious and somewhat complicated algebraic manipulations (see Gross and Harris, for details). We proceed to find the probability Pr $\{N(t) \geqslant n\}$. From (3 5a) we get

$$\begin{aligned}\sum_{r=n}^{\infty} f_r(s) &= \frac{1 - \beta}{s} \sum_{r=n}^{\infty} \beta^r = \frac{1 - \beta}{s} \cdot \frac{\beta^n}{1 - \beta} = \frac{\beta^n}{s} \\ &= (\lambda/\mu)^n\, \alpha^{-n}/s, \qquad\qquad (3.6) \\ &\equiv L\{A_n(t)\}, \; n \geqslant 0,\end{aligned}$$

where $\quad A_n(t) \equiv \Pr\{N(t) \geqslant n\} = p_n(t) + p_{n+1}(t) + \dots$

The function $A_n(t)$ can be obtained by finding the inverse L.T. of (3.6). From the tables of L.T., we find that the L.T. of $\left\{\dfrac{n}{t}\, I_n(at)\right\}$ is

$$\left\{\frac{a}{s + \sqrt{(s^2 - a^2)}}\right\}^n.$$

Using the translation property, the L.T. of

$$a_n(t) \equiv \left(\frac{\lambda}{\mu}\right)^{n/2} e^{-(\lambda+\mu)t} \left\{\frac{n}{t}\ I_n(2t\sqrt{\lambda\mu})\right\} \tag{3.7}$$

is found to be

$$\left(\frac{\lambda}{\mu}\right)^{n/2} \left\{\frac{2\sqrt{(\lambda\mu)}}{(s+\lambda+\mu)+\sqrt{\{(s+\lambda+\mu)^2-4\lambda\mu\}}}\right\}^n = \left(\frac{\lambda}{\mu}\right)^n \alpha^{-n}$$

and hence the L.T. of $\int_0^t a_n(u)\,du$

is $$(\lambda/\mu)^n\, \alpha^{-n}/s.$$

Thus the derivative of $A_n(t)$ is $a_n(t)$, given by (3.7) (A_n is the distribution with density a_n).

The steady state probability $p_n = \lim_{t\to\infty} p_n(t)$ can be obtained by taking limit of $p_n(t)$ obtained after inverting $f_n(s)$. It can also be obtained easily by using the initial value theorem of Laplace transform.

Now $$\lim_{s\to 0} \beta \equiv \lim_{s\to 0} \beta(s) = \lim_{s\to 0} \frac{(s+\lambda+\mu)-\sqrt{\{(s+\lambda+\mu)^2-4\lambda\mu\}}}{2\mu}$$

$$= \begin{cases} \dfrac{(\lambda+\mu)-(\mu-\lambda)}{2\mu} = \rho, \text{ when } \lambda < \mu \\[2ex] \dfrac{(\lambda+\mu)-(\lambda-\mu)}{2\mu} = 1, \text{ when } \lambda \geqslant \mu. \end{cases} \tag{3.8}$$

Thus $$p_n = \lim_{t\to\infty} p_n(t) = \lim_{s\to 0} sf_n(s)$$

$$= \lim_{s\to 0} (1-\beta)\,\beta^n$$

$$= \begin{cases} (1-\rho)\rho^n, & \rho < 1 \\ 0 & , \rho \geqslant 1. \end{cases}$$

The result $p_n = 0$, $\rho \geqslant 1$ may be interpreted by saying that when the traffic intensity is greater than or equal to 1, there is ultimately zero probability that the system contains a finite number of units. One would intuitively expect that the queue size would grow larger and larger in such cases.

Another fact to be noted is that, when $\rho = 1$, the states are persistent and null having infinite recurrence time.

Method of Generating Functions

To cover the more general case, let us suppose that $N(0) = i$, i.e. the process starts with i units at time $t = 0$; and we denote

$$\Pr\{N(t) = n \mid N(0) = i\} \text{ by } p_{in}(t) \text{ so that } p_{ij}(0) = \delta_{ij}$$

(so far we assumed that $N(0) = 0$; putting $i = 0$ and writing $p_n(t)$ for $p_{in}(t)$ when $i = 0$ we can get back to the case last studied). For the difference equations (2.1) and (2.2) we shall now get

$$p'_{i0}(t) = -\lambda p_{i0}(t) + \mu p_{i1}(t) \tag{3.9}$$

$$p'_{in}(t) = -(\lambda + \mu)\, p_{in}(t) + \lambda p_{in-1}(t) + \mu p_{in+1}(t), \; n = 1, 2, \ldots \tag{3.10}$$

Let

$$P(z, t) = \sum_{n=0}^{\infty} p_{in}(t)\, z^n$$

be the generating function of $\{p_{in}(t)\}$. Then $P(z, 0) = z^i$. Let

$$f_{in}(s) \equiv L\{p_{in}(t)\}$$

and

$$F(z, s) \equiv L\{P(z, t)\}.$$

Multiplying (3.10) by z^n and adding for $n = 1, 2, \ldots,$ the resulting equations with (3.9) we get

$$\sum_{n=0}^{\infty} p'_{in}(t)\, z^n = -(\lambda + \mu) \sum_{n=0}^{\infty} p_{in}(t)\, z^n + \mu p_{i0}(t)$$

$$+ \lambda \sum_{n=1}^{\infty} p_{in-1}(t)\, z^n + \mu \sum_{n=0}^{\infty} p_{in+1}(t)\, z^n$$

or $$\frac{\partial}{\partial t} P(z, t) = -(\lambda + \mu)\, P(z, t) + \mu p_{i0}(t) + \lambda z P(z, t)$$

$$+ (\mu/z)\{P(z, t) - p_{i0}(t)\}$$

or $$z \frac{\partial}{\partial t} P(z, t) = \{\lambda z^2 - (\lambda + \mu)\, z + \mu\}\, P(z, t) - \mu\,(1 - z)\, p_{i0}(t). \tag{3.11}$$

Taking L.T. and noting that

$$L\left\{\frac{\partial}{\partial t} P(z, t)\right\} = sF(z, s) - P(z, 0)$$

$$= sF(z, s) - z^i,$$

we get

$$szF(z, s) - z^{i+1} = \{\lambda z^2 - (\lambda + \mu)\, z + \mu\}\, F(z, s) - \mu(1 - z)\, f_{i0}(s)$$

or

$$F(z, s) = \frac{\mu(1 - z)\, f_{i0}(s) - z^{i+1}}{\lambda z^2 - (\lambda + \mu + s)\, z + \mu}. \tag{3.12}$$

This involves $f_{i0}(s)$ which can be evaluated as follows. The denominator of the r.h.s. has two zeros, say, ξ and η,

where

$$\xi, \eta \equiv \frac{(s + \lambda + \mu) \pm K}{2\lambda};$$

further, $|\xi| < 1$, $|\eta| > 1$. Note that the roots of the equation $\lambda z^2 - (\lambda + \mu + s)\, z + \mu = 0$ are the reciprocals of the roots of the equation (3.4). Thus ξ, η are the reciprocals of α, β, i.e. $\xi = 1/\alpha$, $\eta = 1/\beta$, and thus $|\xi| < 1$ and $|\eta| > 1$. It can be shown that $\lambda z^2 - (\lambda + \mu + s)\, z + \mu = 0$ has only one zero $z = \xi$ in the unit circle. Since $F(z, s)$ converges in the region $|z| = 1$, the zeros in the unit circle of the numerator and the denominator of the r.h.s. of (3.12) must coincide, so that $z = \xi$ must be a zero of the numerator also. Hence

$$f_{i0}(s) = \xi^{i+1}/\mu(1 - \xi)$$

and finally

$$F(z, s) = \frac{(1-z)\,\xi^{i+1} - z^{i+1}(1-\xi)}{\lambda(1-\xi)(z-\xi)(z-\eta)}. \tag{3.13}$$

Expanding the r.h.s. of (3.13) in powers of z, $f_{in}(s)$ can be obtained as the coefficient of z^n.

Particular case: When $i = 0$, i.e., $N(0) = 0$, we get

$$F(z, s) = \frac{1}{\lambda(1-\xi)(\eta - z)} = \frac{1}{\lambda(1 - 1/\alpha)(1/\beta - z)} \tag{3.14}$$

$$= \frac{\alpha\beta}{\lambda(\alpha - 1)} \sum_{n=0}^{\infty} \beta^n z^n = \frac{(1-\beta)}{s} \sum_{n=0}^{\infty} \beta^n z^n$$

$$(\text{since } \mu(\alpha-1)(\beta-1) = -s).$$

Hence,

$$f_n(s) = f_{0n}(s)$$

$$\equiv \text{coeff. of } z^n \text{ in } F(z, s)$$

$$= \frac{(1-\beta)\,\beta^n}{s}, \text{ as in (3.5a).}$$

Note: That $|\alpha| > 1$ and $|\beta| < 1$ can be best demonstrated by applying: *Rouche's Theorem.* If $f(z)$ and $g(z)$ are functions analytic inside and on a closed contour C, and if $|g(z)| < |f(z)|$ on C, then $f(z)$ and $f(z) + g(z)$ have the same number of zero (s) inside C.

(For a proof of this theorem, reference may be made to any standard book on Complex-variable theory.)

Here we take C to be $|z| = 1$, $f(z) = (\lambda + \mu + s)\,z$, $g(z) = \lambda + \mu z^2$, and Re $(s) > 0$. We have, on $|z| = 1$,

$$|g(z)| = |\lambda + \mu z^2| \leqslant \lambda + \mu < |s + \lambda + \mu| = |(s + \lambda + \mu)\,z| = |f(z)|.$$

Now $f(z) = (s + \lambda + \mu)\,z$ has only one zero inside $|z| = 1$ and so $\mu z^2 + \lambda - (s + \lambda + \mu)\,z = 0$ also has only one root inside $|z| = 1$. This equation has two roots α, β of which β has the smaller modulus. Thus $|\beta| < 1$ and $|\alpha| > 1$.

A more sophisticated approach not involving generating function and Rouche's theorem has been recently given by Neuts (1979).

Busy Period : *Zero-avoiding state probabilities*

A busy period is defined as the interval of time commencing at the instant 0 when a unit arrives at an empty counter and terminating at the instant when the server becomes free for the first time. Let the length of the interval which is a random variable be T, and $\{N^*(t)\}$ be the stochastic process denoting the number of units present at the instant t during the busy period. We have $N^*(0) = 1$ and the duration of the busy period is the first t for which $N^*(t) = 0$. Let

$$q_n(t) = \Pr\{N^*(t) = n \mid N^*(0) = 1\}, \tag{3.15}$$

be the *zero-avoiding state probabilities*.

Then $q_1(0) = 1$, $q_n(0) = 0, n \geqslant 2$. For $n \geqslant 2$, $q_n(t)$ will satisfy the same differential equations as $p_n(t)$, i.e. the equations (2.2) will hold also for $q_n(t)$ for $n = 2, 3, \ldots$, i.e.

$$q_n'(t) = -(\lambda + \mu)\, q_n(t) + \lambda q_{n-1}(t) + \mu q_{n+1}(t). \tag{3.16}$$

The equations corresponding to $n = 1$ and $n = 0$ will be as follows:

$$q_1'(t) = -(\lambda + \mu)\, q_1(t) + \mu q_2(t) \tag{3.17}$$

$$q_0'(t) = \mu q_1(t). \tag{3.18}$$

Let $g_n(s)$ be the L.T. of $q_n(t)$. Then taking L.T. of (3.16) we get, for $n \geqslant 2$,

$$sg_n(s) - q_n(0) = -(\lambda + \mu)\, g_n(s) + \lambda g_{n-1}(s) + \mu g_{n+1}(s)$$

or
$$\mu g_{n+1}(s) - (s + \lambda + \mu)\, g_n(s) + \lambda g_{n-1}(s) = 0.$$

The solution of this equation is (as in (3.3))

$$g_n(s) = A\alpha^n + B\beta^n, \quad n \geqslant 2, \ |\alpha| > 1, \ |\beta| < 1.$$

We can choose the constants A and B so that they yield correct values of $g_1(s)$ and $g_0(s)$. Again, $\sum\limits_n g_n(s)$ must converge so that A is equal to zero. Thus putting $n = 1$, we get $g_1(s) = B\beta$ so that

$$g_n(s) = g_1(s)\, \beta^{n-1}, \quad n \geqslant 1. \tag{3.19}$$

Taking L.T. of (3.17) we find

$$s\, g_1(s) - q_1(0) = -(\lambda + \mu)\, g_1(s) + \mu\beta\, g_1(s),$$

whence
$$g_1(s) = \frac{1}{s + \lambda + \mu - \mu\beta} = \frac{\beta}{\lambda} \quad \text{(as } \beta \text{ satisfies (3.4))}.$$

Thus
$$g_n(s) = g_1(s)\, \beta^{n-1} = \beta^n/\lambda = \rho^n/\lambda\alpha^n, \quad n = 1, 2, \ldots \tag{3.20}$$

Inverting the L.T., we get, for $n = 1, 2, \ldots$

$$q_n(t) = \frac{n\rho^{(n/2)}}{\lambda t}\, e^{-(\lambda+\mu)t}\, I_n(2t\sqrt{\lambda\mu}). \tag{3.21}$$

Let $b(t)$ be the p.d.f. of the busy period T. Then, for small δt,

$$b(t)\, \delta t = \Pr\{t \leqslant T < t + \delta t\}$$

equals the probability that the busy period terminates in $(t, t + \delta t)$. It is clear that this probability equals the probability that there are j ($\geqslant 1$) units in the system at the instant t and that the services of these j units are completed in $(t, t + \delta t)$. Thus

$$b(t)\, \delta t = \sum_{j=1}^{\infty} [\, \Pr\{N^*(t) = j \mid N^*(0) = 1\}$$
$$\cdot \Pr\{j \text{ units complete service in } (t, t + \delta t)\}]$$

$$= q_1(t)\, \mu\, \delta t + \sum_{j=2}^{\infty} q_j(t)\, (o(\delta t))$$

so that

$$\begin{aligned} b(t) &= \mu\, q_1(t) \\ &= \frac{1}{t}\,\rho^{-\frac{1}{2}}\, e^{-(\lambda+\mu)t}\, I_1(2t\sqrt{\lambda\mu}). \end{aligned} \tag{3.22}$$

(using (3.21)).

The L.T. of the p.d.f. of T is

$$b^*(s) = L\{b(t)\} = L\{\mu\, q_1(t)\} = \mu\, g_1(s) = \beta/\rho. \tag{3.23}$$

Using (3.8) we get,

$$\begin{aligned} \Pr\{T < \infty\} = \lim_{s\to 0+} b^*(s) &= 1, \quad \text{if} \quad \rho < 1 \\ &= 1/\rho, \quad \text{if} \quad \rho \geqslant 1, \end{aligned}$$

indicating that there is non-zero probability that the busy period lasts for ever when $\rho \geqslant 1$.

We now find the moments of T. We have for $\rho < 1$

$$\frac{d}{ds}\beta(s) = \frac{1}{2\mu}\{1 - 2(s+\lambda+\mu)/2K\} = -\frac{\beta}{K}, \ (K \equiv K(s))$$

$$\frac{d^2}{ds^2}\beta(s) = \frac{\beta}{K^2} + \frac{\beta}{K^3}(s+\lambda+\mu) = \frac{2\lambda}{K^3}.$$

Thus

$$E(T) = -\frac{d}{ds}b^*(s)\Big|_{s=0} = \frac{\beta}{\rho K}\Big|_{s=0} = \frac{1}{\mu-\lambda} = \frac{1}{\mu(1-\rho)}, \tag{3.24}$$

$$E(T^2) = \frac{d^2}{ds^2}b^*(s)\Big|_{s=0} = \frac{2\lambda}{\rho}\cdot\frac{1}{(\mu-\lambda)^3} = \frac{2}{\mu^2(1-\rho)^3} \tag{3.25}$$

and

$$\operatorname{var}(T) = \frac{1+\rho}{\mu^2(1-\rho)^3}.$$

The coefficient of variation $C_V = \dfrac{(\text{var } T)^{\frac{1}{2}}}{E(T)} \times 100$

$$= \frac{(1+\rho)^{\frac{1}{2}}}{(1-\rho)^{\frac{1}{2}}} \times 100$$

increases as ρ increases, while it is constant for W_S. The following table gives $E(T)$ and var (T) for some values of λ, μ ($\mu = 1$, $\lambda = \rho$).

$\lambda\ (=\rho)$	0.2	0.4	0.5	0.8	0.9	0.95	0.995
$E(T)$	1.25	1.67	2	5	10	20	200
var(T)	2.34	6.48	12	225	1900	15600	1596×10^4

It is clear that C_V is large for large ρ, i.e. var(T) is large compared to $E(T)$; as such the expected value does not seem to have much significance for large ρ.

Note: (1) *Idle* (or *free period*) is the duration of the interval from the instant the server becomes free for the first time to the instant of the next arrival when he becomes engaged. This duration being the residual portion of the interarrival time is again exponential with the same mean $1/\lambda$, since the interarrival distribution is exponential with mean $1/\lambda$.

(2) We have considered the initial busy period, i.e. the period initiated by one unit. The distribution of the busy period γ_j initiated by j units is given by the convolution $\tau_j = \sum_{i=1}^{j} T_i$, where T_i are i.i.d. random variables distributed as T considered above.

(3) The concept of busy period is due the French mathematician Emil Borel (1871–1956), who obtained the joint bivariate distribution of T and N, the number served during the busy period T for $M/D/1$ model. Various approaches are available for obtaining the distribution of T as well as the joint distribution of T and N in more general cases Distributions of T and N can be obtained as marginal distributions of the joint distribution. The trivariate distribution of T, N and M, where M denotes the maximum queue length attained during the busy period T, has also been engaging attention of researchers.

Example 3(a). *First passage time distribution.* For a $M/M/1$ queue in transient state, let τ be the random variable denoting the time taken for the system to *fall* from the initial state a to another state $b(<a)$ for the first time. Let $X(t)$ be the process denoting such states of the system, given $X(0)=a$,

$$q_n(t) = \Pr\{X(t) = n \mid X(0) = a\}. \tag{3.26}$$

Using similar arguments it can be easily shown that the state probabilities satisfy the following equations:

$$\begin{aligned} q_n'(t) &= -(\lambda+\mu)\, q_n(t) + \lambda q_{n-1}(t) + \mu q_{n+1}(t),\ n \geqslant (b+2) \\ q_{b+1}'(t) &= -(\lambda+\mu)\, q_{b+1}(t) + \mu q_{b+2}(t) \\ q_b'(t) &= \mu q_{b+1}(t). \end{aligned} \tag{3.27}$$

The p.d.f. $f(t)$ of the first passage time distribution from a to b is given by $f(t) = \mu q_{b+1}(t) = q_b'(t)$.

Putting $a=1$, $b=0$, we can get back to initial busy period distribution.

Waiting time process. Let $W(t)$ be the time required to serve all the units present in the system at the instant t, given that $W(0)=0$. Then, if $N(t)$ is the number present at instant t, $(N(0)=0)$,

$$W(t) = \begin{cases} 0, & \text{if } N(t) = 0 \\ v_1' + v_2 + \ldots + v_{N(t)}, & \text{if } N(t) > 0 \end{cases}$$

where v_1' is the residual service time of the unit being served at the instant t, and $v_2, \ldots, v_{N(t)}$ those of the waiting units at the instant t. $\{W(t), t \geqslant 0\}$, which is known as *virtual waiting time*, is a Markov process (with continuous state space). Given $W(0)=0$, then proceeding as in the case of waiting time in the system in steady state it can be seen that its probability element $f(x,t)dx = \Pr\{x \leqslant W(t) < x + dx\}$, $0 \leqslant x < \infty$, $0 \leqslant t < \infty$ is given by

$$f(x,t)\,dx = \sum_{n=0}^{\infty} \left\{\mu \frac{(\mu x)^n}{\Gamma(n+1)} e^{-\mu x}\, dx\right\} p_{n+1}(t).$$

Its L.T. can be put in elegant closed forms (see Prabhu (1965)).

10.4 Birth and Death Processes in Queueing Theory: Multichannel Models

A. Birth and Death Processes

Let us first consider a birth and death process with state dependent birth and death rates λ_n and μ_n respectively. Let $N(t)$ be the number present at the instant t, and

$$p_n(t) = \Pr\{N(t) = n \mid N(0) = \cdot\}$$

It was shown in Ch. 4 that $\{N(t), t \geqslant 0\}$ is a Markov process with denumerable state space $\{0,1,2,\ldots\}$ and that the forward Kolmogorov equations of the process are:

$$p_0'(t) = -\lambda_0 p_0(t) + \mu_1 p_1(t)$$

$$p_n'(t) = -(\lambda_n + \mu_n)\, p_n(t) + \lambda_{n-1} p_{n-1}(t) + \mu_{n+1} p_{n+1}(t), \quad n = 1, 2, \ldots$$

(see equations (4.4) and (4.6), Ch. 4).

We proceed to investigate the steady state solutions. Assume that such solutions exist, then $\lim\limits_{t\to\infty} p_n(t) = p_n = \Pr\{N = n\}$, N being the random variable giving the number of units. Putting $p_n(t) = p_n$ and $p_n'(t) = 0$, we get from the above the following difference equations in steady state:

$$0 = -\lambda_0 p_0 + \mu_1 p_1 \tag{4.1}$$

$$0 = -(\lambda_n + \mu_n)\, p_n + \lambda_{n-1} p_{n-1} + \mu_{n+1} p_{n+1}, \quad n = 1, 2, \ldots. \tag{4.2}$$

From (4.2), we have for $n = 1, 2, \ldots,$

$$\begin{aligned} \mu_{n+1} p_{n+1} - \lambda_n p_n &= \mu_n p_n - \lambda_{n-1} p_{n-1} \\ &= \mu_{n-1} p_{n-1} - \lambda_{n-2} p_{n-2} \quad \text{(putting } n-1 \text{ for } n) \\ &\quad \ldots \qquad\qquad \ldots \\ &= \mu_1 p_1 - \lambda_0 p_0 \\ &= 0 \qquad \text{(from (4.1))}, \end{aligned}$$

so that

$$\begin{aligned} p_{n+1} &= \frac{\lambda_n}{\mu_{n+1}}\, p_n \\ &= \left(\frac{\lambda_n}{\mu_{n+1}}\right)\left(\frac{\lambda_{n-1}}{\mu_n}\right) p_{n-1} \\ &= \frac{\lambda_n \lambda_{n-1} \ldots \lambda_0}{\mu_{n+1} \mu_n \ldots \mu_1}\, p_0 \end{aligned}$$

or

$$p_n = \prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k}\, p_0, \quad n = 1, 2, \ldots. \tag{4.3}$$

Since $\sum\limits_{n=0}^{\infty} p_n$ must be unity,

$$\left\{1 + \sum_{n=1}^{\infty} \prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k}\right\} p_0 = 1. \tag{4.4}$$

Hence a necessary and sufficient condition for the existence of a steady state is the convergence of the infinite series $\sum_{n=1}^{\infty}\prod_{k=1}^{n}\frac{\lambda_{k-1}}{\mu_k}$ (Karlin and McGregor (1957)). When it converges, p_0 can be found from (4.4).

The queueing processes arising in some of the standard models can be considered as birth and death processes and the steady state solutions can be obtained easily by using the above. In the $M/M/1$ model considered in Sec. 10.2, $\lambda_n = \lambda, n = 0, 1, 2, \ldots, \mu_n = \mu, n = 1, 2, \ldots$. Putting these values in (4.3) and (4.4) we at once get the steady state solutions.

B. The Model M/M/s

Here we consider a queueing model with $s\,(1 \leqslant s \leqslant \infty)$ servers or channels in parallel and having identical input and service time distributions as in the model $(M/M/1)$. In other words, the present model $(M/M/s)$ considers a Poisson process with parameter λ as its input process and has, for each of the s channels, i.i d. exponential service time distribution with mean rate μ. If $n\,(\leqslant s)$ channels are busy, the number of services completed in the whole system is a Poisson process with mean $n\mu$ and thus the time between two successive service completions is exponential with mean $1/n\mu$; whereas if $n\,(> s)$ channels are busy, the time between two successive service completions is exponential with mean $1/s\mu$. If $N(t)$ is the number present in the system at the instant t, the transition densities are as follows:

$$\begin{aligned} a_{n,n+1} &\equiv \Pr\{N(t+dt) = n+1 \mid N(t) = n\} = \lambda dt + o(dt)\\ a_{n,n-1} &\equiv \Pr\{N(t+dt) = n-1 \mid N(t) = n\} = \mu_n dt + o(dt)\\ a_{n,m} &\equiv \Pr\{N(t+dt) = m \mid N(t) = n\} = o(dt),\ m \neq n-1, n+1 \end{aligned}$$

where

$$\begin{aligned} \mu_n &= n\mu, \quad \text{if} \quad 0 \leqslant n \leqslant s\\ &= s\mu \quad \text{if} \quad n \geqslant s. \end{aligned} \tag{4.5}$$

Thus $N(t)$ is a birth and death process with constant arrival (birth) rate $\lambda_n = \lambda$ and state dependent service (death) rates as given in (4.5).

Let $p_n(t) = \Pr\{N(t) = n \mid N(0) = \cdot\}$, and let the steady state solutions exist.

Putting the values of λ_n and μ_n in (4.3) and (4.4) we get p_0 and p_n, $n = 1, 2, \ldots$, as follows:

$$p_n = \frac{\lambda\cdot\lambda\ldots\lambda}{(\mu)(2\mu)\ldots(n\mu)}p_0 = \frac{(\lambda/\mu)^n}{n!}p_0,\ n = 1, 2, \ldots, s \tag{4.6}$$

$$\text{and } p_n = \frac{\lambda\cdot\lambda\ldots n \text{ factors}}{\{(\mu)\}(2\mu)\ldots(s\mu)\}\{(s\mu)\cdot(s\mu)\ldots(n-s)\text{ factors}\}}p_0$$

$$= \frac{\lambda^n}{s!\,\mu^s s^{n-s}\mu^{n-s}}p_0 = \frac{(\lambda/\mu)^n}{s!\,s^{n-s}}p_0, \quad n \geqslant s.$$

The condition $\sum_{n=0}^{\infty} p_n = 1$ gives

$$p_0^{-1} = 1 + \sum_{n=1}^{s-1}\frac{(\lambda/\mu)^n}{n!} + \sum_{n=s}^{\infty}\frac{(\lambda/\mu)^n}{s!\,s^{n-s}} = \sum_{n=0}^{s-1}\frac{(\lambda/\mu)^n}{n!} + \frac{s^s}{s!}\sum_{n=s}^{\infty}\left(\frac{\lambda}{s\mu}\right)^n.$$

For existence of steady state solutions, the series $\sum_{n=s}^{\infty}\left(\frac{\lambda}{s\mu}\right)^n$ must converge, and for this to happen $\lambda/s\mu$ must be less than 1. Then

$$p_0^{-1} = \sum_{n=0}^{s-1}\frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!\,(1-\lambda/s\mu)} \tag{4.7}$$

Thus the distribution of N, when $\rho = \lambda/s\mu < 1$, is given by (4.6), where p_0 is given by (4.7).

Note: (1) In this s-channel model, $\rho = \lambda/s\mu$ is the traffic intensity.

(2) When $s = 1$, we get the corresponding result for $M/M/1$.

(3) The solutions p_n satisfy the following recurrence relations

$$\begin{aligned} p_n &= \frac{1}{n}\left(\frac{\lambda}{\mu}\right) p_{n-1},\ n = 1, 2, \ldots, s-1 \\ &= \frac{1}{s}\left(\frac{\lambda}{\mu}\right) p_{n-1},\ n = s, s+1, \ldots \end{aligned}$$

(4) The probability that an arriving unit has to wait is given by

$$\begin{aligned} \Pr\{N \geqslant s\} &= \sum_{n=s}^{\infty} p_n \\ &= \frac{(\lambda/\mu)^s}{s!\,(1-\rho)}\, p_0 = \frac{p_s}{1-\rho}\,. \end{aligned} \tag{4.8}$$

(5) The expected number in the system $E\{N\} = \sum_{n=0}^{\infty} np_n$ can be directly evaluated by finding the sum of the series. Likewise $E\{Q\}$, the expected number in the queue can be found (see Prabhu, Gross and Harris). Here we shall obtain them by making use of Little's formula.

Waiting Time Distributions

We assume that the queue discipline is FCFS. The distribution of W_Q can be obtained in the same way as in $M/M/1$.

Consider a test unit. It has not to wait, if it finds, on its arrival, less than or equal to $(s-1)$ units in the system. Thus

$$\begin{aligned} \Pr\{W_Q = 0\} &= \sum_{n=0}^{s-1} p_n \\ &= 1 - \sum_{n=s}^{\infty} p_n \\ &= 1 - \frac{(\lambda/\mu)^s}{s!\,(1-\rho)}\, p_0 = 1 - \frac{p_s}{1-\rho} \quad \text{(using (4.8)).} \end{aligned}$$

The test unit has to wait, i.e. $W_Q > 0$ when it finds, on arrival, that the number in the system is $n \geqslant s$, of which s are in the service channels and $(n-s)$ in the queue. Thus it will have to wait for the completion of $(n-s+1)$ services. As $n \geqslant s$, s-channels are always in operation before its service commences, and as such the time between two successive service completions is exponential with mean $1/s\mu$. Thus

S_{n-s+1}, being the sum of $(n-s+1)$ i.i.d. exponential variates with mean $1/s\mu$, is a gamma variate with p.d.f.

$$\frac{(s\mu)(s\mu x)^{n-s}\, e^{-s\mu x}}{(n-s)!}$$

Hence
$$\begin{aligned} w_q(x)\,dx &= \Pr\{x \leqslant W_Q \leqslant x+dx\}, x>0 \\ &= \sum_{n=s}^{\infty} \Pr\{x < S_{n-s+1} < x+dx \mid N=n\}\cdot \Pr\{N=n\} \\ &= \sum_{n=s}^{\infty} \frac{(s\mu)\,(s\mu x)^{n-s}}{(n-s)!}\, e^{-s\mu x}\, p_n\, dx \end{aligned}$$

or
$$\begin{aligned} w_q(x) &= \frac{(s\mu)e^{-s\mu x}}{s!}\, p_0 \sum_{n=s}^{\infty} \frac{(s\mu x)^{n-s}}{(n-s)!}\, \frac{r^n}{s^{n-s}},\ (r=\lambda/\mu) \\ &= \frac{(s\mu)e^{-s\mu x}}{s!}\, p_0 r^s \sum_{m=0}^{\infty} \frac{(r\mu x)^m}{m!},\ (\text{putting } n-s=m) \\ &= \frac{s\mu\,(\lambda/\mu)^s}{s!}\, e^{-\mu x\{s-(\lambda/\mu)\}}\, p_0. \\ &= s\mu\, e^{-x(s\mu-\lambda)}\, \frac{(\lambda/\mu)^s}{s!}\, p_0 \\ &= s\mu\, e^{-x(s\mu-\lambda)}\, p_s. \end{aligned}$$

Thus
$$\begin{aligned} w_q(x) &= 1 - \frac{(\lambda/\mu)^s}{(s-1)!(s-\lambda/\mu)}\, p_0,\quad x=0 \\ &= \frac{(\lambda/\mu)^s}{(s-1)!}\, \mu e^{-(s\mu-\lambda)x}\, p_0,\quad x>0 \end{aligned} \tag{4.9}$$

and $E\{W_Q\} = (W_Q = 0)\cdot\Pr\{W_Q=0\} + \int_{\varepsilon}^{\infty} x w_q(x)\,dx$ as $\varepsilon \to 0$.

Now
$$\begin{aligned} \int_{\varepsilon}^{\infty} x w_q(x)\,dx &= \int_{\varepsilon}^{\infty} \frac{(\lambda/\mu)^s}{(s-1)!}\, p_0\mu\,\{x e^{-(s\mu-\lambda)x}\}\,dx \\ &\to \left\{\frac{(\lambda/\mu)^s}{(s-1)!}\,\mu p_0\right\} \frac{1}{(s\mu-\lambda)^2} = \frac{(\lambda/\mu)^s}{s\mu(s!)(1-\rho)^2}\, p_0. \end{aligned}$$

Thus
$$E\{W_Q\} = \frac{(\lambda/\mu)^s}{s\mu(s!)(1-\rho)^2}\, p_0 = \frac{p_s}{s\mu(1-\rho)^2}. \tag{4.10}$$

Corollary. The expected waiting time in the system is given by

$$E\{W_S\} = E\{W_Q\} + 1/\mu.$$

Using Little's formula, we get

$$\begin{aligned} E\{L_Q\} = \lambda E\{W_Q\} &= \frac{\rho(\lambda/\mu)^s}{s!\,(1-\rho)^2}\, p_0 \\ &= \frac{\rho}{(1-\rho)^2}\, p_s. \end{aligned} \tag{4.11}$$

and
$$E\{N\} = \lambda E\{W_S\} = \frac{\lambda}{\mu} + \frac{\rho(\lambda/\mu)^s}{s!\,(1-\rho)^2}\, p_0. \tag{4.12}$$

If the random variables, number of busy and idle channels, are denoted by B and I respectively, then $N = Q + B$ and $I + B = s$,

whence we get $$E\{B\} = E\{N\} - E\{Q\} = \lambda/\mu$$

and $$E\{I\} = s - \lambda/\mu = s(1 - \rho).$$

Note: Suppose that instead of having *homogeneous servers* (viz. all with i.i.d. exponential distribution with rate μ) *heterogeneous servers* having independent exponential service time with rates μ_i, $i = 1, 2, \ldots, s$, are considered. It can be seen that the waiting time density is, for $x > 0$, given by

$$\omega_q(x) = [\{\sum_{i=1}^{s} \mu_i\} \exp\{-x(\sum_{i=1}^{s} \mu_i - \lambda)\}]\, p_s$$

(Gumbel).

Example 4(a). Patients arrive at the outpatient department of a hospital in accordance with a Poisson process at the mean rate of 12 per hour, and the distribution of time for examination by an attending physician is exponential with a mean of 10 minutes. What is the minimum number of physicians to be posted for ensuring a steady state distribution? For this number, find (i) the expected waiting time of a patient prior to being examined and (ii) the expected number of patients in the out-patient department. How many physicians, on an average, will remain idle?
Here $\lambda = 12$/hr., $\mu = 6$/hr., so that $r = \lambda/\mu = 2$. For steady state, $\rho = r/s < 1$, i e. $2 < s$, the minimum number of physicians must be 3. We take $s = 3$, then $\rho = 2/3$. From Exercise 10.7, we find that

$$E\{W_Q\} = \left[s\mu(1-\rho)\left\{1 + (1-\rho)\frac{E(s, r)}{p(s, r)}\right\}\right]^{-1}$$

where $$p(s, r) = \frac{e^{-r} r^s}{s!}, \quad E(\ , r) = \sum_{k=0}^{s-1} \frac{e^{-r} r^k}{k!}.$$

From Poisson probability ta. . $p(3, 2) = 0.180$, $E(3, 2) = 0.677$. Extensive tables of Poisson probability function and distribution function are available (Molina (1942)).

We get $$E\{W_Q\} = \frac{2}{27}\text{ hr} = 4.4\text{ min.}$$

$$E\{W_S\} = E\{W_Q\} + 1/\mu = 13/54\text{ hr} = 14.4\text{ min.}$$

$$E\{N\} = \lambda E\{W_S\} = 26/9$$

and $$E\{I\} = s(1 - \rho) = 3(1 - 2/3) = 1.$$

C. The Model M/M/∞

Here the number of channels or servers are infinite so that, on arrival, all the units are taken into service and there is no queue. For the birth and death process associated with this, $\lambda_n = \lambda$, $\mu_n = n\mu$, for $n = 0, 1, 2, \ldots$. The steady state solutions can be obtained from those of $M/M/s$ by letting s tend to infinity. From (4.6) and (4.7) we at once get the steady state probabilities,

$$p_n = \frac{(\lambda/\mu)^n}{n!} p_0, \quad n = 1, 2, \ldots,$$

where p_0 is obtained by using $\sum\limits_{n=0}^{\infty} p_n = 1$; we get

$$p_0^{-1} = \sum_{n=0}^{\infty} \frac{(\lambda/\mu)^n}{n!} = e^{\lambda.\mu}.$$

Thus
$$p_n = \frac{(\lambda/\mu)^n}{n!} e^{-\lambda/\mu}, \; n = 0, 1, 2, \ldots. \tag{4.13}$$

The distribution of the number N in the system is Poisson with mean λ/μ. The distribution of the number of busy channels is also the same.

As the series $\sum\limits_{n=0}^{\infty} (\lambda/\mu)^n/n!$ converges for all λ/μ, the steady state solutions exist whether or not $\lambda < \mu$.

D. The Model M/M/s/s : Loss system

This model envisages that a unit, who finds, on arrival, that all the s-channels are busy leaves the system without waiting for service. This is called a (s-channel) *loss system* and was first investigated by Erlang. This model was also examined earlier (see Examples 5(d) and 5(f), Ch. 4).

For this birth and death process, we have

$$\begin{aligned} \lambda_n &= \lambda, \quad & \mu_n &= n\mu, \; n = 0, 1, 2, \ldots, s-1 \\ &= 0, & &= s\mu, \; n \geqslant s. \end{aligned} \tag{4.14}$$

From (4.3) and (4.4) we get the steady state probabilities

$$p_n = \frac{(\lambda/\mu)^n}{n!} p_0, \quad n = 0, 1, \ldots, s$$

and
$$\left\{1 + \sum_{k=1}^{s} \frac{(\lambda/\mu)^k}{k!}\right\} p_0 = 1 \quad \text{or} \quad p_0 = \left[\sum_{k=0}^{s} \frac{(\lambda/\mu)^k}{k!}\right]^{-1}$$

Thus
$$p_n = \frac{(\lambda/\mu)^n/n!}{\sum\limits_{k=0}^{s} \dfrac{(\lambda/\mu)^k}{k!}}, \quad n = 0, 1, 2, \ldots, s. \tag{4.15}$$

The probability that an arriving unit is lost to the system (which is the same as that all the the channels are busy) is given by

$$p_s = \frac{(\lambda/\mu)^s/s!}{\sum\limits_{k=0}^{s} (\lambda/\mu)^k/k!}. \tag{4.16}$$

The formula (4.16) is known as *Erlang's loss formula* or *blocking formula* and is denoted by $B(s, \lambda/\mu)$, while (4.15) is known as Erlang's first formula (or simply Erlang's formula, the corresponding distribution being truncated Poisson).

Note: (1) Attempts have been made since Erlang's time to generalise Erlang's results. Mention may be made of the works of Pollaczek, Palm, Kosten, Fortet, Sevast'yanov and Takacs (cf. his (1969) paper). It has been shown that Erlang's formula (4.16) holds for *any* distribution of service time (having mean $1/\mu$) provided the input is Poisson (with parameter λ), i.e. it holds for the model $M/G/s/s$ (loss system).

(See also Brumelle (1978) for a generalisation to state dependent arrival and service rates.)

If $N(t)$ is the number of busy channels at instant t, then $\{N(t), t \geqslant 0\}$ is a Markov process, provided the service time is exponential. But the process is not Markov when the service time is not exponential; nevertheless, the loss formula holds.

(2) The steady state solutions p_n exist whether or not $\lambda < \mu$.

(3) The results of the model $M/M/\infty$ can be obtained by taking limits, as $s \to \infty$, of the corresponding results for $M/M/s/s$.

Example 4(b). *System with limited waiting time.* Apart from limitation of space, a second kind limitation which arises in problems dealing with practical situations needs consideration. For example, a long distance telephone call may be booked to be put through within a limited time (say, during the office hours). If the call does not mature by that time, the call is cancelled and is lost. Similarly, perishable goods can often be stored for a limited time. In these cases, if a unit does not receive service, immediately on arrival, the unit waits but waits only for a limited time, on the expiry of which it is lost.

We assume that the system has Poisson input with rate λ, has s-channels with i.i.d. exponential service times each with mean $1/\mu$, that there is no limitation as regards waiting space. Further, we assume that the waiting time of a unit is a random variable having exponential distribution with mean $1/\nu$.

The resulting process is Markov with state space $\{0, 1, 2, \ldots\}$, and can be described by a birth and death process. Here $\lambda_n = \lambda$ for $n = 0, 1, \ldots,$ $\mu_n = n\mu$ for $n = 0, 1, \ldots, s$. For $n > s$, the transition from state n to state $(n-1)$ in an infinitesimal interval h may occur in two mutually exclusive ways: either one of the s busy servers terminates his service (with probability $s\mu h + o(h)$) or the waiting time of one of the $(n-s)$ units waiting in the queue expires (with probability $(n-s)\nu h + o(h)$). Thus

$$\mu_n = s\mu + (n-s)\nu,\ n = s+1, s+2, \ldots$$

Putting the values of λ_n, μ_n in (4.3), we get

$$\begin{aligned} p_n &= \frac{(\lambda/\mu)^n}{n!} p_0, \quad n \leqslant s \\ &= \frac{\lambda^n p_0}{s!\,\mu^s \prod\limits_{=s+1}^{n} \{s\mu + (r-s)\nu\}}, \quad n > s \end{aligned}$$

and using $\sum_{n=0}^{\infty} p_n = 1$, we get

$$p_0 = \left[\left\{ \sum_{k=1}^{s} \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k \frac{\lambda^n}{\mu^s (s!)} \right\} \sum_{n=s+1}^{\infty} \frac{\lambda^n}{\prod_{r=s+1}^{n} \{s\mu + (r-s)\nu\}} \right]^{-1}.$$

E. The Model with Finite Input Source

We have assumed so far that the source or the population from which the input comes is infinite. We now consider the case where the source of input is not infinite but is finite and consists of m units. For example, such a situation occurs in practice in servicing of automatic machines. Suppose that there are m such machines, which normally do not need human attention except when there is a breakdown. In case of failure or breakdown of a machine, it is serviced by one of the $s\,(< m)$ repairmen available. If all the s repairmen are engaged in repairing machines already down, the machine that is down at that time waits for repair till a repairman becomes free. We have a finite input source conisting of m machines and $s\,(< m)$ repairmen.

Suppose that, if a machine is in working state at the instant t, the probability that it will breakdown in the infinitesimal interval $(t, t+h)$ is $\lambda h + o(h)$ and that, when a machine is being serviced, the probability that the servicing is completed in $(t, t+h)$ is $\mu h + o(h)$. Let $N(t)$ be the number of machines not working (or down for repair) at the instant t. We have a birth and death process with parameters

$$\begin{aligned} \lambda_n &= (m-n)\lambda, & n &= 0, 1, 2, \ldots, m-1 \\ &= 0, & n &\geqslant m. \end{aligned}$$

and

$$\begin{aligned} \mu_n &= n\mu, & n &= 0, 1, 2, \ldots, s-1 \\ &= s\mu, & n &\geqslant s. \end{aligned} \tag{4.17}$$

Using (4.3), we get in steady state,

$$\begin{aligned} p_n &= \frac{(m)(m-1)(m-2)\ldots(m-n+1)\lambda^n}{n!\,\mu^n}\, p_0 \\ &= \frac{m!}{(m-n)!\,n!} \left(\frac{\lambda}{\mu}\right)^n p_0 \\ &= \binom{m}{n} \left(\frac{\lambda}{\mu}\right)^n p_0, \text{ for } n = 0, 1, 2, \ldots, s-1 \end{aligned} \tag{4.18}$$

and

$$p_n = \frac{m!}{(m-n)!\,s!\,s^{n-s}} \left(\frac{\lambda}{\mu}\right)^n p_0, \text{ for } n = s, s+1, \ldots, m; \tag{4.19}$$

where p_0 is obtained by using $\sum_{n=0}^{m} p_n = 1$; we get

$$p_0 = \left[\sum_{n=0}^{s-1} \binom{m}{n} \left(\frac{\lambda}{\mu}\right)^n + \sum_{n=s}^{m} \frac{m!}{(m-n)!\,s!\,s^{n-s}} \left(\frac{\lambda}{\mu}\right)^n \right]^{-1}. \tag{4.20}$$

The probability that the number of machines in working order is r ($\leqslant m$) is given by p_{m-r}.

The model has been used as an answer to various important questions of practical interest arising out of *machine interference problems*. Some such questions relate to determination of the optimal number of repairmen for a given set of machines, the maximum length of time during which all the m machines are working and so on. These have been the subject matter of investigation of several researchers (e.g. Palm, Naor, Benson and Cox; see also Feller (1968) p. 465).

In a paper, Levhari and Shesinski (1970) used such a model in a somewhat broader perspective: in the study of microeconomic production function. On the assumption that the machines produce a constant flow of output when operating, and that the breakdown and servicing of machines conform to the model of machine interference studied, they obtain the steady state probabilities of the system as well as the relation between inputs and output and fit the same to a Cobb-Douglas production function.

Remarks: In Sec. 10.4A we have considered generalised birth and death processes, with state dependent mean arrival and service rates λ_n and μ_n On the basis of of the results indicated there, queueing models with *special choices* of λ_n and μ_n have been analysed in subsequent subsections. Recently Conolly (1975), Conolly and Chan (1977) consider generalised birth and death queueing process, with state dependent rates λ_n and μ_n, as they, are. They assume the condition for the existence of steady state, and define a measure of traffic intensity in terms of 'effective' service time and interarrival interval. The results obtained show surprising similarity (with some of those of $M/M/1$ and with Little's formula). The results also indicate potentiality of application to a wider range of systems.

Example 4(c). *Transient solution of $M/M/1/1$ model.* Here $\lambda_0 = \lambda, \mu_0 = \mu$ and $\lambda_1 = 0$, $\mu_n = \mu$, $n \geqslant 1$ (see eqn. (4.14)): if $N(t)$ denotes the number in the system at time t, then $\Pr\{N(t) = n\} = p_n(t) = 0$ for all $n > 1$, i.e. we are concerned with only $p_0(t)$ and $p_1(t)$ such that $p_0(t) + p_1(t) = 1$. The forward Kolmogorov equations of the model then become

$$p_0'(t) = -\lambda p_0(t) + \mu p_1(t)$$

$$p_1'(t) = -\mu p_1(t) + \lambda p_0(t).$$

Writing $p_1(t) = 1 - p_0(t)$ in the first equation we get $p_0'(t) + (\lambda + \mu)\, p_0(t) = \mu$. The solution of this first order linear differential equation with constant coefficients is given by

$$p_0(t) = Ce^{-(\lambda+\mu)t} + \frac{\mu}{\lambda + \mu}, \text{ where } C \text{ is a constant.}$$

Given the initial distribution $p_i(0) = \Pr\{N(0) = i\}$, we get

$$p_0(t) = p_0(0)\, e^{-(\lambda+\mu)t} + \frac{\mu}{\lambda + \mu}\, \{1 - e^{-(\lambda+\mu)t}\}.$$

Similarly, $p_1(t) = p_1(0)\, e^{-(\lambda+\mu)t} + \dfrac{\lambda}{\lambda+\mu}\{1 - e^{-(\lambda+\mu)t}\}$.

The steady state solutions are

$$p_0 = \lim_{t\to\infty} p_0(t) = \mu/(\lambda+\mu)$$

$$p_1 = \lim_{t\to\infty} p_1(t) = \lambda/(\lambda+\mu),$$

irrespective of whether the value of $\rho = \lambda/\mu < 1$ or not. (These agree with $p_n(n = 0, 1)$ given by (4.15).)

Assume that the initial distribution is identical with the steady-state distribution so that $p_0(0) = p_0 = \mu/(\lambda+\mu)$, $p_1(0) = p_1 = \lambda/(\lambda+\mu)$. Then we find that *for all* $t \geqslant 0$,

$$p_0(t) = \frac{\mu}{\lambda+\mu} = p_0,\ p_1(t) = \frac{\lambda}{\lambda+\mu} = p_1.$$

That is, if the process is in equilibrium (steady state) initially, then it will be always (for all $t \geqslant 0$) in steady state. This is true for any other '*ergodic*' system.

In case of $M/M/s/s$ model, if the system is in equilibrium initially, i.e. $\Pr\{N(0) = n\} = p_n(0) = p_n$, $0 \leqslant n \leqslant s$, where p_n are steady-state probabilities (given by relation (4.15)), then $\{N(t), t \geqslant 0\}$ becomes a stationary process for which $N(t)$ has the same distribution for all $t \geqslant 0$, i.e. $p_n(t) = \Pr\{N(t) = n\} = p_n$ (given by (4.15)) *for all* $t \geqslant 0$ (see Takacs (1969)).

10.5 Non-birth and Death Queueing Processes: Bulk Queues

We have seen that the queueing processes in the standard models considered in the earlier sections are birth and death processes and are, of course, Markovian. We examine here queueing processes arising out of some non-birth and death processes arising out of some non-birth and death processes but which are nevertheless Markovian. The Chapman-Kolmogorov equations can be written and the analysis can be done in a similar manner.

The basic model considered is of the type of a $M/M/1$ queue with parameters λ, μ, but with group arrivals and/or batch service.

A. Group Arrival System: $M^{(X)}/M/1$

Assume that the arrival streams form a Poisson process with parameter λ and that the number of units arriving at an arrival instant is a random variable X with p.f. $c_k = \Pr\{X = k\}$, $k = 1, 2, \ldots$. This is equivalent to saying that, in an infinitesimal interval $(t, t+h)$, the probability that a group of k arrives is $\lambda c_k h + o(h)$. Assume here that service takes place singly and that the distributions of the service times are independent exponential with mean $1/\mu$. It is clear that the process is still Markovian but is an non-birth—death process, as transitions may occur to states, not necessarily neighbouring. We denote the model by $M^{(X)}/M/1$. We get

$$p_0'(t) = -\lambda p_0(t) + \mu p_1(t) \tag{5.1}$$

$$p_n'(t) = -(\lambda+\mu)\,p_n(t) + \mu p_{n+1}(t) + \sum_{k=1}^{n} \lambda c_k\, p_{n-k}(t),\ n \geqslant 1. \tag{5.2}$$

Let $C(s) = \sum_{k=1}^{\infty} c_k s^k$ be the p.g.f. of X. Since the mean arrival rate is $\lambda\, E(X) = \lambda C'(1)$, the traffic intensity can be defined as

$$\rho = \lambda E(X)/\mu = \lambda C'(1)/\mu. \tag{5.3}$$

Assuming that the steady state solutions exist, we get

$$0 = -\lambda p_0 + \mu p_1 \tag{5.4}$$

$$0 = -(\lambda+\mu)\,p_n + \mu p_{n+1} + \lambda \Sigma\, c_k p_{n-k},\ n \geqslant 1. \tag{5.5}$$

Let $P(s) = \sum_{n=0}^{\infty} p_n\, s^n$ be the p.g.f. of $\{p_n\}$.

Multiplying (5.5) by s^n, for successive values of $n = 1, 2, 3, \ldots$ and adding them to (5.4), we get

$$0 = -\lambda\, P(s) - \mu\,(P(s) - p_0) + (\mu/s)\,(P(s) - p_0) + \lambda \sum_{n=1}^{\infty} \sum_{k=1}^{n} c_k\, p_{n-k}\, s^n. \tag{5.6}$$

Noting that the last term equals $\lambda \sum_{k=1}^{\infty} c_k\, s^k \{\sum_{n=k}^{\infty} p_{n-k}\, s^{n-k}\} = \lambda C(s)\, P(s)$, we get

$$P(s) = \frac{\mu\,(1-s)\,p_0}{\mu\,(1-s) - \lambda s\,(1 - C(s))}. \tag{5.7}$$

To find p_0, we use $P(1) = 1$. Applying l'hopital's rule, we get

$$1 = \lim_{s \to 1} P(s) = \frac{-\mu p_0}{-\mu + \lambda C'(1)}$$

or

$$p_0 = 1 - \rho. \tag{5.8}$$

Thus

$$P(s) = \frac{\mu\,(1-s)\,(1-\rho)}{\mu\,(1-s) - \lambda s\,(1 - C(s))}. \tag{5.7a}$$

Particular case: $M/M/1$. This model corresponds to single arrivals, i.e. in groups of size one only. Then $c_1 = 1$, $c_k = 0$, $k > 1$ and

$$P(s) = (1-\rho)/(1-\rho s).$$

B. Bulk Service Systems

Bulk Service Rules

There could be a number of a policies or rules according to which batches for bulk service may be formed. The following are the types of bulk service policies or rules frequently discussed in the literature.

(1) Bailey (1954) (also Downton, 1955) considers that units are served in batches of not more than b (say). If, immediately after the completion of a service, the server finds more than b units waiting, he takes a batch of b for service while others wait; if he finds r units $(0 \leqslant r \leqslant b)$, he takes all the r units in a batch for service. Bloemena (1960), Jaiswal (1960) Neuts (1967) consider the same rule with the restriction that $r \neq 0$ $(1 \leqslant r \leqslant b)$, i.e. the service facility stops until a unit arrives. This rule will be called the *usual bulk service rule*, while Bailey's rule will be its *modified* type (it will be also called *bulk service rule with intermittently available server*). Jaiswal points out that the distribution of the queue length for the modified rule can be obtained from that of the usual rule.

(2) The rule with a fixed batch size k has been considered by Fabens (1961) Takacs (1962) and others. In this case the server waits until there are k units.

(3) Neuts (1967) considers this rule: If, immediately after the completion of a service, the server finds less than a units present, he waits until there are a units, whereupon he takes the batch of a units for service; if he finds a or more but at most b, he takes them all in the batch and if he finds more than b, he takes in the batch for service b units, while others wait. The batch takes a minimum of a units and a maximum of b units. This rule will be called *general bulk service rule* as rules under (1) and (2) can be taken as particular cases.

(4) Bhat (1963) considers the rule that the number taken in a batch is a random variable Y. The corresponding Markov model is denoted by $M/M^{(Y)}/1$.

There are actual situations where one rule seems to be better suited or to be more appropriate than the others. Examples are given in standard textbooks and relevant research publications.

Poisson Queue with General Bulk Service Rule: $M/M^{a,b}/1$*

Assume that the arrivals occur singly in accordance with a Poisson process with parameter λ. The units are served in batches according to a general bulk service rule such that the service starts only when a minimum number of units, say, a is present in the queue, the maximum service capacity being, say, b. The service times of batches of size s $(a \leqslant s \leqslant b)$ are assumed to have independent identical exponential distribution with mean $1/\mu$. For simplicity, the service time distribution is assumed to be independent of the batch size s. Denote by state $(0, q)$, the state that the service channel is idle and there are q $(0 \leqslant q \leqslant a-1)$ units waiting in the queue and by the state $(1, n)$, the state that the service channel is busy and there are n $(n \geqslant 0)$ units waiting in the queue. Let

$$p_{0,q}(t) = \Pr\{\text{that at time } t\text{, the system is in the state } (0, q)\},$$
$$q = 0, 1, 2, \ldots, a-1,$$

---

*The queueing system is denoted here by $M/M^{a,b}/1$.

$p_{1,n}(t) = \Pr$ {that at time t, the system is in the state $(1, n)$},
$n = 0, 1, 2, \ldots$

Assume that the steady state distribution exists and let

$$\lim_{t\to\infty} p_{0,q}(t) = p_{0,q}$$

$$\lim_{t\to\infty} p_{1,n}(t) = p_{1,n}.$$

The steady state probability equations are :

$$0 = -(\lambda + \mu)\, p_{1,n} + \lambda p_{1,n-1} + \mu p_{1,n+b}, \quad n = 1, 2, \ldots, \tag{5.9}$$

$$0 = -(\lambda + \mu)\, p_{1,0} + \lambda p_{0,a-1} + \mu \sum_{s=a}^{b} p_{1,s} \tag{5.10}$$

$$0 = -\lambda\, p_{0,0} + \mu\, p_{1,0} \tag{5.11}$$

$$0 = -\lambda p_{0,n} + \lambda p_{0,n-1} + \mu p_{1,n}, \quad n = 1, 2, \ldots, a-1. \tag{5.12}$$

Note that (5.12) will not occur when $a = 1$. First we solve (5.9).

The difference equation (5.9) can be written as

$$h(E)\{p_{1,n}\} = 0, \quad n = 0, 1, 2, \ldots \tag{5.13}$$

such that the characteristic equation is

$$h(z) \equiv \mu z^{b+1} - (\lambda + \mu)\, z + \lambda = 0. \tag{5.14}$$

Suppose that $\quad f(z) \equiv -(\lambda + \mu)\, z \quad$ and that $g(z) \equiv \mu z^{b+1} + \lambda$,

then for $\quad |z| = 1$, $|g(z)| < |f(z)|$ and hence by Rouché's theorem, $f(z)$ and $f(z) + g(z)$ will have the same number of zeros inside $|z| = 1$. Since $f(z)$ has only one zero inside $|z| = 1$, there will be only one zero of $f(z) + g(z) = h(z)$ inside $|z| = 1$.

Denote this root of $h(z) = 0$ by r, $0 < r < 1$, and the other b roots by $r_1, \ldots, r_b$, $(|r_i| \geqslant 1)$.

Denote the traffic intensity by $\rho = \lambda/b\mu$. From (5.14), we get

$$b\rho = \frac{\lambda}{\mu} = \frac{r\,(1 - r^b)}{1 - r} = r + r^2 + \ldots + r^b. \tag{5.15}$$

Now we can write the solution of (5.9) as:

$$p_{1,n} = Ar^n + \sum_{i=1}^{b} A_i r_i^{\,n}, \quad n = 0, 1, 2, \ldots$$

Since $\sum_{n=0}^{\infty} p_{1,n} < 1$, we must have $A_i = 0$ for $i = 1, 2, \ldots, b$, so that,

$$\begin{aligned} p_{1,n} &= Ar^n \\ &= (p_{1,0})r^n \\ &= (\lambda/\mu)r^n\, p_{0,0} \quad \text{(from (5.11))} \\ &= \left(\frac{1 - r^b}{1 - r}\right) r^{n+1}\, p_{0,0}. \end{aligned} \tag{5.16}$$

Using (5.16), we get from (5.10)

$$p_{0,\,a-1}=\frac{1-r^{a}}{1-r}\,p_{0,0} \tag{5.17}$$

and finally putting $n=1, 2, \ldots\ a-1$ in (5.12), and using (5.16) and (5.17) we get, recursively for $q=a-2, a-3, \ldots, 2, 1$,

$$p_{0,q}=\frac{1-r^{q+1}}{1-r}\,p_{0,0} \tag{5.18}$$

and because of (5.17), (5.18) holds for $q=1, 2, \ldots, a-1$.

Since $\sum\limits_{q=0}^{a-1} p_{0,q}+\sum\limits_{n=0}^{\infty} p_{1,n}=1$, we have

$$p_{0,0}=\left[\frac{a}{1-r}+\frac{r^{a+1}-r^{b+1}}{(1-r)^2}\right]^{-1}. \tag{5.19}$$

Thus (5.16), (5.18), (5.19) give the steady state probabilities.

Particular case : Fixed size bulk service : $M/M^{k,k}/1$.

Here $a=b=k$. Let p_n denote the steady state probability that the number in the *system* is n. Then

$$p_0=p_{0,0}=\frac{1-r}{k},$$

$$p_q=p_{0,q}=\frac{1-r^{q+1}}{1-r}p_0=(1-r^{q+1})/k,\ q=1,2,\ldots,k-1,$$

$$p_{m+k}=p_{1,m}=(\lambda/\mu)\,p_0\,r^m=\rho\,(1-r)\,r^m,\ \rho=\lambda/(k\mu),$$
$$m=0,1,2,\ldots,$$

i.e.

$$p_n=\begin{cases}(1-r^{n+1})/k, & n=0,1,2,\ldots,k-1\\ \rho\,(1-r)\,r^{n-k}, & n=k,k+1,\ldots\end{cases} \tag{5.20}$$

Distribution of the waiting time for the system $M/M^{a,b}/1$ (Medhi, 1975)

Assume that the system is in steady state. Let the random variable T denote the waiting time in the queue for an arriving unit.

Let

$v(t)$ be the p.d.f. of T,

$f(\alpha, k; t)$ be the p.d.f. of gamma distribution with parameters α, k, i.e.

$$f(\alpha, k; t)=\alpha^k t^{k-1}\,e^{(-\alpha t)}/\Gamma(k),\ t>0,\ k=1,2,\ldots,$$

$\Gamma_x(\alpha, k)$ be the incomplete gamma function, i.e.

$$\Gamma_x(\alpha, k)=\int_0^x f(\alpha, k; t)\,dt$$

$$=1-\sum_{s=0}^{k-1} e^{-\alpha x}\,(\alpha x)^s/s!.$$

An arriving unit may find the system in one of the following states:

$$\begin{aligned}
&\text{(i)} \quad (0, a-1)\\
&\text{(ii)} \quad (0, q),\ 0 \leqslant q \leqslant a-2\\
&\left.\begin{aligned}&\text{(iii)} \quad (1, n), a-1 \leqslant m \leqslant b-1\\ &\text{(iv)} \quad (1, n), 0 \leqslant m \leqslant a-2\end{aligned}\right\} \quad \begin{aligned}&n = kb + m\\ &k = 0, 1, 2, \ldots\end{aligned}
\end{aligned} \tag{5.21}$$

In case of (i), the arriving unit does not wait; the probability of zero delay is thus $p_{0,a-1}$. In all other cases the unit has to wait, the probability of blocking (or delay) is thus $1 - p_{0,a-1}$.

In case of (ii), the arriving unit has to wait the arrival of $(a-1-q)$ units and the time needed for $(a-1-q)$ arrivals has a gamma distribution with parameters λ, $a-1-q$.

In case of (iii), the arriving unit has to wait for the completion of services of $(k+1)$ batches; the time required for this has a gamma distribution with parameters μ, $k+1$.

In case of (iv), the arriving unit has to wait till either the services of $(k+1)$ groups are completed or $(a-1-m)$ units arrive, whichever occurs later. This duration which is given by the maximum of two gamma variates may be denoted by the random variable Z, i.e.

$$\begin{aligned} Z = \max\{&\text{gamma variate with parameters } \lambda, a-1-m;\\ &\text{gamma variate with parameters } \mu, k+1\}. \end{aligned} \tag{5.22}$$

The distribution function $F_Z(t)$ and the p.d.f. $h_Z(t)$ of Z are given by

$$F_Z(t) = \Pr\{Z \leqslant t\} = \Gamma_t(\lambda, a-1-m)\,\Gamma_t(\mu, k+1),$$

and $h_Z(t) = F'_Z(t)$

$$\begin{aligned} &= f(\lambda, a-1-m; t)\,\Gamma_t(\mu, k+1)\\ &\qquad + \Gamma_t(\lambda, a-1-m)\,f(\mu, k+1; t). \end{aligned} \tag{5.23}$$

It follows that the p.d.f. $v(t)$ of T is given by

$$\begin{aligned} v(t) = &\sum_{q=0}^{a-2} p_{0,q}\, f(\lambda, a-1-q; t)\\ &+ \sum_{k=0}^{\infty} \sum_{m=a-1}^{b-1} p_{1,\, kb+m}\, f(\mu, k+1; t).\\ &+ \sum_{k=0}^{\infty} \sum_{m=0}^{a-2} p_{1,\, kb+m}\, h_Z(t), \quad 0 < t < \infty. \end{aligned} \tag{5.24}$$

It may be verified that

$$\int_0^\infty v(t)\, dt = 1 - p_{0,\, a-1},$$

as is to be expected.

Substituting the values of $p_{0,q}$, $p_{1,n}$ and the expressions for f and h in (5.24), we get an exact expression for the p.d.f. $v(t)$ as follows: Denote

$$\sum_{k=0}^{m-1} z^k/k! \equiv e(m, z);\ e^{-z}.\ e(m, z) \equiv E(m, z),\ m \geqslant 1. \tag{5.25}$$

Then

$$\sum_{q=0}^{a-2} r^q f(\lambda, a-q-1; t) = \lambda r^{a-2} \exp(-\lambda t)\, e(a-1, \lambda t/r).$$

Thus the first member on the right hand side (r.h.s) of (5.24) becomes

$$\lambda p_{0,0}\, [E(a-1, \lambda t) - r^{a-1} \exp(-\lambda t)\, e(a-1, \lambda t/r)]/(1-r). \qquad (5.26)$$

The second member on the r.h.s. of (5.24) is equal to

$$(\lambda/\mu)\, p_{0,0}\, [\{\sum_{m=a-1}^{b-1} r^m\} \{\mu \exp(-\mu t) \sum_{k=0}^{\infty} (\mu t r^b)^k/k!\}]$$
$$= [\lambda p_{0,0}\, (r^{a-1} - r^b) \exp\{-\mu(1-r^b)t\}]/(1-r). \qquad (5.27)$$

The third member on the r.h.s. of (5.24) consists of two terms corresponding to those of (5.23); the first of these becomes

$$(\lambda/\mu)\, p_{0,0} \left[\left\{\sum_{m=0}^{a-2} r^m f(\lambda, a-1-m; t)\right\}\left\{\int_0^t \sum_{k=0}^{\infty} (r^b)^k f(\mu, k+1; x)\, dx\right\}\right]$$
$$= \lambda p_{0,0}\, [r^{a-1} \exp(-\lambda t)\, e(a-1, \lambda t/r)]$$
$$\cdot [1 - \exp\{-\mu(1-r^b)\, t\}]/(1-r), \qquad (5.28)$$

and the second becomes

$$(\lambda/\mu) p_{0,0}\, [\sum_{k=0}^{\infty} (r^b)^k f(\mu, k+1; t)][\sum_{m=0}^{a-2} r^m\, \Gamma_t(\lambda, a-1-m)]$$
$$= \lambda p_{0,0}\, [\exp\{-\mu(1-r^b)\, t\}] \left[\sum_{m=0}^{a-2} r^m \left\{1 - \sum_{q=0}^{a-2-m} \frac{e^{-\lambda t}(\lambda t)^q}{q!}\right\}\right]$$
$$= \lambda p_{0,0}\, [\exp\{-\mu(1-r^b)\, t\}] \left[\frac{1-r^{a-1}}{1-r} - \frac{E(a-1, \lambda t)}{1-r} + \frac{r^{a-1} \exp(-\lambda t)\, e(a-1, \lambda t/r)}{1-r}\right]. \qquad (5.29)$$

Hence adding the expressions of the r.h.s of (5.26) to (5.29), we get

$$v(t) = [\lambda p_{0,0}/(1-r)][(1-r^b) \exp\{-\mu(1-r^b)t\}$$
$$+ E(a-1, \lambda t)\{1 - \exp(-\mu(1-r^b)t)\}], \; 0 < t < \infty. \quad (5.30)$$

The p.d.f. of the conditional waiting time distribution, given that a customer has to wait (i.e. an arriving unit who finds the service channel busy or idle with less than $a-1$ waiting) is given by

$$w(t) = v(t)/\{1 - p_{0,a-1}\}. \qquad (5.31)$$

Moments of the distribution of T

To find the moments of T, we note that for positive integral values of k and for $a \geqslant 2$,

$$J_1(k) \equiv \int_0^{\infty} t^k \exp\{-\mu(1-r^b)\, t\}\, dt = \frac{\Gamma(k+1)}{\{\mu(1-r^b)\}^{k+1}}, \qquad (5.32)$$

$$J_2(k) \equiv \int_0^\infty t^k E(a-1, \lambda t)\, dt$$

$$= \sum_{s=0}^{a-2} \int_0^\infty t^k (\lambda t)^s e^{-\lambda t}\, dt/s\,!$$

$$= \sum_{s=0}^{a-2} (s+1)(s+2)\ldots(s+k)/\lambda^{k+1}, \tag{5.33}$$

$$J_3(k) \equiv \int_0^\infty t^k E(a-1, \lambda t) \exp\{-\mu(1-r^b)\, t\}\, dt$$

$$= \sum_{s=0}^{a-2} (s+1)(s+2)\ldots(s+k)\, r^{s+k+1}/\lambda^{k+1}. \tag{5.34}$$

The expected waiting time in the queue $E(T)$ is given by

$$E(T) = \int_0^\infty t v(t)\, dt = \frac{\lambda p_{0,0}}{1-r} \left[(1-r^b)\, J_1(1) + J_2(1) - J_3(1)\right]$$

$$= \frac{\lambda p_{0,0}}{1-r} \left[\frac{1}{\mu^2(1-r^b)} + \frac{a(a-1)}{2\lambda^2} + \frac{ar^{a+1}(1-r) - r^2(1-r^a)}{\lambda^2(1-r)^2}\right]. \tag{5.35}$$

It can be verified that Little's formula holds.
Higher moments can be obtained in the same manner.

Particular case: $M/M^{1,b}/1$. For the usual bulk service rule,

$$p_{0,0} = \frac{(1-r)}{1-r+(\lambda/\mu)} = \frac{(1-r)^2}{(1-r)^2 + r(1-r^b)},$$

$$p_{1,n} = \frac{(1-r)(1-r^b)}{(1-r)^2 + r(1-r^b)}\, r^{n+1}, \quad n = 0, 1, 2, \ldots .$$

The waiting time density is given by

$$v(t) = [\lambda p_{0,0}/(1-r)][(1-r^b) \exp\{-\mu(1-r^b)\, t\}], \tag{5.36}$$

and the expected waiting time by

$$E(T) = \frac{r}{\mu[(1-r)^2 + r(1-r^b)]}. \tag{5.37}$$

The conditional waiting time distribution of only those who wait is exponential with mean $1/\mu\{1-r^b\}$.

Putting $b = 1$ (when $r = \rho$), we get the corresponding results for $M/M/1$.

Remarks: Subsequent to Neuts (1967), who introduced the general bulk service rule for single channel systems, such models have been studied by Teghem *et al.* (1969), Medhi *et al.* (1972), Medhi (1975, 1979), Borthakur (1975, 1976), Neuts (1979), Sim, Templeton *et al.*; Arora (1964), Ghare (1968), Cromie *et al.* (1976), Chaudhry *et al.*, Weiss (1979), and others examine some similar bulk service models.

10.6 The Queueing Models $M/G/1$ and $GI/M/1$

So long we have been discussing queueing processes which are either birth and death or non-birth and death processes. They are in either case Markovian and the theory of Markov chains and processes could be applied in their studies. We shall now consider models where the distributions of the interarrival time or the service time do not possess the memoryless property, i.e. are not exponential. The process $N(t)$ giving the state of the system or system size at time t will then be no longer Markovian; however, the analysis of the process can be based on an associated process which is Markovian. Two techniques are generally being used for this purpose. Kendall (1951) uses the concept of regeneration point (due to Palm) by suitable choice of regeneration points and extracts, from the process $\{N(t)\}$, Markov chains in discrete time at those points. This is known as the technique of imbedded Markov chains. The second important technique (due to Keilson and Kooharian (1960)) known as supplementary variable technique, involves inclusion of such variable(s).

We discuss below Kendall's method.

A. Queues with Poisson Input: model $M/G/1$

Assume that the input process is Poisson with intensity λ and that the service times are i.i.d. random variables having an arbitrary (general) distribution with mean $1/\mu$. Denote the service time by v, its d.f., by $B(t)$, its p.d.f., when it exists, by $b(t)(= B'(t))$, and its L.T. by

$$B^*(s) = \int_0^\infty e^{-st}\, dB(t).$$

Let $t_n, n = 1, 2, \ldots, (t_0 = 0)$ be the nth departure epoch, i.e. the instant at which the nth unit completes his service and leaves the system. These points t_n are the regeneration points of the process $\{N(t)\}$. The sequence of points $\{t_n\}$ forms a renewal process. $N(t_n + 0)$, the number in the system immediately after the nth departure has a denumerable state space $\{0, 1, 2, \ldots\}$. Write $N(t_n + 0) \equiv X_n$, $n = 0, 1, \ldots$ and denote by A_n, the random variable giving the number of units that arrive during the service time of the nth unit. Then

$$\begin{aligned} X_{n+1} &= X_n - 1 + A_{n+1}, && \text{if } X_n \geqslant 1 \\ &= A_{n+1}, && \text{if } X_n = 0. \end{aligned} \tag{6.1}$$

Now the service times of all the units have the same distribution so that $A_n \equiv A$ for $n = 1, 2, \ldots$. We have

$$\Pr\{A = r \mid \text{service time of a unit is } t\} = \frac{e^{-\lambda t}(\lambda t)^r}{r!}$$

and so

$$k_r \equiv \Pr\{A = r\} = \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^r}{r!}\, dB(t),\ r = 0, 1, 2, \ldots, \tag{6.2}$$

gives the distribution of A, the number of arrivals during the service time of a unit. The probabilities

$$p_{ij} = \Pr\{X_{n+1} = j \mid X_n = i\}$$

are given by

$$\begin{aligned} p_{ij} &\equiv k_{j-i+1}, \ i \geqslant 1, j \geqslant i-1 \\ &= 0 \quad , i \geqslant 1, j < i-1 \\ p_{0j} &\equiv p_{1j} = k_j, \ j \geqslant 0. \end{aligned} \tag{6.3}$$

The relations (6.3) clearly indicate that $\{X_n, n \geqslant 0\}$ is a Markov chain having t.p.m.

$$P = (p_{ij}) = \begin{pmatrix} k_0 & k_1 & k_2 & \dots \\ k_0 & k_1 & k_2 & \dots \\ 0 & k_0 & k_1 & \dots \\ \dots & & \dots & \dots \\ \dots & & \dots & \dots \end{pmatrix}. \tag{6.4}$$

As every state can be reached from every other state, the Markov chain $\{X_n\}$ is irreducible. Again as $p_{ii} \geqslant 0$, the chain is aperiodic. It can also be shown that, when the traffic intensity $\rho = \lambda/\mu < 1$, the chain is persistent, non-null and hence ergodic. We can then apply the ergodic theorem of Markov chain (Theorem 3.6).

The limiting probabilities

$$v_j = \lim_{n\to\infty} p_{ij}^{(n)}, \ j = 0, 1, 2, \dots$$

exist and are independent of the initial state i. The probabilities $V = (v_0, v_1, \dots)$, $\Sigma v_j = 1$, are given as the unique solutions of

$$V = VP.$$

Let $K(s) = \Sigma k_j s^j$ and $V(s) = \Sigma v_j s^j$ denote the p.g.f. of the distributions of $\{k_j\}$ and $\{v_j\}$ respectively.

We have $K(s) = \sum\limits_{j=0}^{\infty} k_j s^j = \sum\limits_{j=0}^{\infty} s^j \left\{ \int\limits_0^\infty \frac{e^{-\lambda t}(\lambda t)^j}{j!} dB(t) \right\}$

$$= \int_0^\infty e^{-(\lambda-\lambda s)t} \, dB(t)$$

$$= B^*(\lambda - \lambda s). \tag{6.5}$$

Hence

$$E(A) = K'(1) = -\lambda B^{*(1)}(0) = \lambda/\mu = \rho. \tag{6.6}$$

Now $V = VP$ gives an infinite system of equations.

Multiplying the $(k+1)$st equation by s^k, $k = 0, 1, \dots$ and adding over k, we get, on simplification, for $0 < \rho < 1$,

$$V(s) = \frac{\{1 - K'(1)\}(1-s)K(s)}{K(s) - s}$$

(see Example 5(d), Ch. 3)

Putting $K'(1) = \rho$, we get

$$V(s) = \frac{(1-\rho)(1-s)\,K(s)}{K(s)-s} \tag{6.7}$$

$$= \frac{(1-\rho)(1-s)\,B^*(\lambda-\lambda s)}{B^*(\lambda-\lambda s)-s}. \tag{6.7a}$$

This is known as *Pollaczek-Khinchine (P.K.) formula.*

Remarks: For the model $M/G/1$, the process $\{N(t)\}$ giving the system size at time t is not a Markov process. By extracting a process at departure points t_n, we get an associated process $X_n = N(t_n + 0)$, which is a Markov chain. The transition probabilities, p_{ij} of the Markov chain, give departure point probabilities, while

$$p_{in}(t) = \Pr\{N(t) = n \mid N(0) = i\}$$

give the general time probabilities of the system size. When the Markov chain is ergodic (the condition for which is $\rho < 1$), the limiting probabilities $v_n = \lim p_{in}^{(m)}$, as $m \to \infty$, exist and can be obtained by applying the ergodic theorem of Markov chain. When the steady state exists (the condition for which is also $\rho < 1$) the limiting probabilities $p_n = \lim p_{in}(t)$, as $t \to \infty$ exist. It can be shown that (for proof, see Cohen, Gross and Harris) when $\rho < 1$, then $v_n = p_n$ for all n, i.e. the limiting departure point probabilities are equal to the corresponding steady state probabilities of the system size. Thus ergodicity of the Markov chain is equivalent to the existence of steady state, the condition for each to hold good being $\rho < 1$.

It may also be noted that another interesting stochastic process can be extracted. The process $Y(t) = X_n$, $t_n \leqslant t < t_{n+1}$ gives the number of units left behind in the system by the most recent departure. Clearly, $\{X_n, t_n\}$ is a Markov renewal process; and $\{Y(t), t \geqslant 0\}$ is a semi-Markov process, X_n being its embedded Markov chain and the waiting time in a state being distributed as the service time v. We shall consider such semi-Markov processes later.

Example 6(a). *The Model $M/E_k/1$.* Consider that the service time v is Erlang-k (with mean $1/\mu$) having p.d.f.

$$dB(t) \equiv b(t) = \frac{(\mu k)^k\, t^{k-1}\, e^{-k\mu t}}{\Gamma(k)},\ 0 < t < \infty.$$

Its L.T. is $$B^*(s) = \left(\frac{\mu k}{s+\mu k}\right)^k \text{ (see Example 3(f), Ch. 1).}$$

From (6.6) and (6.7) we get

$$K(s) = B^*(\lambda - \lambda s) = \left(\frac{\mu k}{\lambda - \lambda s + \mu k}\right)^k$$

$$= \left(1 + \frac{\rho(1-s)}{k}\right)^{-k},\ \rho = \lambda/\mu$$

and
$$V(s) = \frac{(1-\rho)(1-s)}{1 - s\{1 + \rho\,(1-s)/k\}^k}.$$

Expanding in powers of s, and comparing coefficients of s^n from both sides, we get v_n, and equivalently p_n. We can find the moments of $\{p_n\}$ from $V(s)$.

This model could also be considered as non-birth and death Markovian; by considering that the service comprises of k exponential phases each with mean $1/(k\mu)$, the C.K. equations can be written and solutions obtained therefrom.

Special case: *The Model $M/M/1$.* Putting $k = 1$, we at once get (for $M/M/1$),

$$V(s) = \frac{(1-\rho)(1-s)}{(1 - s - s\rho\,(1-s))} = \frac{(1-\rho)}{(1-s\rho)}.$$

Example 6(b). *The model $M/D/1$.* As $k \to \infty$, the whole mass of the distribution E_k tends to concentrate at the point $1/\mu$, i.e. $E_k = D$, as $k \to \infty$. We have then

$$B^*(s) \to e^{-s/\mu},$$

so that

$$K(s) = B^*(\lambda - \lambda s) \to e^{-\rho(1-s)}$$

and

$$V(s) = \frac{(1-\rho)(1-s)}{1 - se^{\rho(1-s)}}.$$

Expanding the r.h.s. in powers of s, and after some algebraic manipulations, we get v_n, the coefficient of s^n, $n = 0, 1, 2, \ldots$, as follows:

$$v_0 = 1 - \rho$$

$$v_1 = (1-\rho)(e^{\rho} - 1)$$

$$\begin{matrix} \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{matrix}$$

$$v_n = (1-\rho)\left[e^{n\rho} + \sum_{r=1}^{n-1}(-1)^{n-r}e^{r\rho}\left\{\frac{(r\rho)^{n-r}}{(n-r)!} + \frac{(r\rho)^{n-r-1}}{(n-r-1)!}\right\}\right],\ n \geqslant 2:$$

WAITING TIME: POLLACZEK-KHINCHINE FORMULA

Assume that the steady state exists ($\rho < 1$).

Let W and W_Q denote the waiting times in the system and in the queue respectively and $W(t)$ and $W_Q(t)$ denote their distribution functions. We have

$$\begin{aligned} p_n = v_n &= \Pr\{\text{departing unit leaves } n \text{ in the system}\} \\ &= \int_0^\infty \Pr\{\text{departing unit leaves } n \text{ in the system} \mid t \leqslant W < t + dt\} \\ &\qquad\qquad . \Pr\{t \leqslant W < t + dt\} \\ &\simeq \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^n}{n!}\, dW(t),\ n = 0, 1, 2, \ldots. \end{aligned} \tag{6.8}$$

Thus

$$\begin{aligned} V(s) &= \sum_{n=0}^{\infty} v_n s^n \\ &= \int_0^{\infty} \{e^{-\lambda t}\, dW(t)\} \left\{ \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} s^n \right\} \\ &= \int_0^{\infty} e^{-\lambda t(1-s)}\, dW(t) \qquad (6.9) \\ &= W^*(\lambda - \lambda s), \end{aligned}$$

where $W^*(\theta) = \int_0^{\infty} e^{-\theta t}\, dW(t)$.

Now,

$$\begin{aligned} P(s) &= \Sigma p_n s^n = \Sigma v_n s^n = V(s) \\ &= W^*(\lambda - \lambda s). \qquad (6.9a) \end{aligned}$$

From (6.7), (6.9) and (6.6) we get

$$W^*(\lambda - \lambda s) = \frac{(1-\rho)(1-s) B^*(\lambda - \lambda s)}{B^*(\lambda - \lambda s) - s}.$$

Writing α for $\lambda - \lambda s$, we get

$$\begin{aligned} W^*(\alpha) &= \frac{(1-\rho)(\alpha/\lambda) B^*(\alpha)}{B^*(\alpha) - (1 - \alpha/\lambda)} \\ &= \frac{\alpha(1-\rho) B^*(\alpha)}{\alpha - \lambda(1 - B^*(\alpha))}, \qquad (6.10) \end{aligned}$$

or,

$$W^*(s) = \frac{s(1-\rho) B^*(s)}{s - \lambda(1 - B^*(s))}. \qquad (6.10a)$$

This relation connects the L.T.'s of the distributions of the service time v and the waiting time W in the system.

Now, $W = W_Q + v$, so that $W^*(s) = W_Q{}^*(s)\, B^*(s)$, $W_Q{}^*$ being the L.T. of W_Q. Thus, we have

$$W_Q{}^*(s) = \frac{s(1-\rho)}{s - \lambda(1 - B^*(s))}. \qquad (6.11)$$

The relation (6.10a) (as well as (6.11)) is also known as Pollaczek-Khinchine (P.K.) formula. To distinguish it from the other P.K. formula (6.7), it is called P.K. *transform formula*. The formula has been rediscovered in special contexts by different researchers. There are many interesting direct proofs of the formula (see also Runnenberg (1960)).

Expected Waiting Time and Expected Number in the System

Using (6.7) and (6.11), moments of $E\{N\}$ and $E\{W_Q\}$ can be obtained.

We have $B^{*(k)}(0) \equiv \frac{d^k}{ds^k} B^*(s) \Big|_{s=0} = (-1)^k E(v^k), k = 1, 2, \ldots.$ (6.12)

From (6.11), we get

$$\frac{W_Q^{*(1)}(s)}{1-\rho} = \frac{s - \lambda + \lambda B^*(s) - s\{1 + \lambda B^{*(1)}(s)\}}{\{s - \lambda + \lambda B^*(s)\}^2},$$

which takes the form $\frac{0}{0}$ as $s \to 0$.

We can use L'Hopital's rule to obtain its value as $s \to 0$. As $s \to 0$, the r.h.s. becomes

$$\frac{(1 + \lambda B^{*(1)}(s)) - (1 + \lambda B^{*(1)}(s)) - s(\lambda B^{*(2)}(s))}{2(s - \lambda + \lambda B^*(s))\,(1 + \lambda B^{*(1)}(s))} = \frac{-\lambda B^{*(2)}(0)}{2\{1 + \lambda B^{*(1)}(0)\}^2}$$

$$= \frac{\lambda E(v^2)}{2(1 - \lambda/\mu)^2}.$$

Thus, $$E(W_Q) = W_Q{}^{*(1)}(0) = \frac{\lambda}{2(1-\rho)} E(v^2)$$

$$= \frac{\lambda}{2(1-\rho)}\left(\sigma_v{}^2 + \frac{1}{\mu^2}\right), \tag{6.13}$$

where $\sigma_v{}^2$ is the variance of the service time v.

By differentiating $W_Q{}^*(s)$, k times w.r.t. s and putting $s = 0$, we can obtain the kth moment of W_Q. It can be seen that

$$\text{var}\,\{W_Q\} = \frac{\lambda}{12(1-\rho)^2}\,[4\,(1-\rho)\,E(v^3) + 3\lambda\,\{E\,(v^2)\}^2] \tag{6.14}$$

It can also be shown that (Sphicas *et al.*, 1978)

$$c^2 \geqslant (4-\rho)/3\rho \geqslant 1,$$

where $c = \text{s.d.}\,\{W_Q\}/E\{W_Q\}$, is the coefficient of variation of W_Q.

Note: By utilising a relationship between single server queues and random walks and without using P.K. formula, Lemoine (1976) obtains the moments of waiting time of $M/G/1$ queue.

The mean waiting time in the system is

$$E\{W\} = E\{W_Q + v\} = E(v) + E\{W_Q\}$$

$$= \frac{1}{\mu} + \frac{\lambda}{2(1-\rho)} E(v^2). \tag{6.15}$$

Using Little's formula, we can obtain $E\{N\}$ from (6.15) and $E\{Q\}$ from (6.13). We have

$$E\{N\} = \lambda E\{W\} = \rho + \frac{\lambda^2}{2(1-\rho)} E(v^2) \tag{6.16}$$

and $$E\{Q\} = \lambda E\{W_Q\} = \frac{\lambda^2}{2(1-\rho)} E(v^2).$$

$E\{N\}, E\{N^2\}, \ldots$ can be obtained directly from (6.7). The formula (6.16) is also sometimes called P.K. formula. Following Kleinrock

(1976) we may call it P.K. *mean value formula*, to distinguish it from P.K. *transform formula*.

From (6.13) and (6.15) it becomes at once clear that for $M/G/1$ queue with parameters λ, μ, expected waiting time, whether in the queue or in the system, is the least when $\sigma_v = 0$, i.e. service time distribution is constant ($= 1/\mu$). The model then becomes $M/D/1$. Thus, for all single server models with Poisson input, the one with constant service time has the least expected waiting time and hence the least expected queue and system size.

We discussed the busy period for the $M/M/1$ model in Sec 10.3; here we consider it for $M/G/1$ model through an alternative approach.

BUSY PERIOD

The expected duration of the busy period T (initiated by a single customer) follows immediately from the result noted in Example 8(a), Ch. 6. For $\rho < 1$, we have

$$p_0 = \lim_{t\to\infty} \{N(t) = 0\} = \frac{E(I)}{E(I) + E(T)}$$

Clearly, the idle period I here is exponential with mean $1/\lambda$, since the interarrival distribution is so. As stated in the remarks noted earlier in this section, $p_n = v_n$ for all n. From equation (6.7) we find that v_0, the constant term in $V(s)$ is given by $v_0 = V(0) = 1 - \rho$, so that

$$p_0 = 1 - \rho = \frac{1/\lambda}{1/\lambda + E(T)}$$

whence $$E(T) = \frac{1}{\mu - \lambda} = \frac{E(v)}{1 - \rho}, \text{ (v being the service time).}$$

Thus, given the mean arrival and service rates, the expected duration of a busy period in a queue with Poisson input is independent of the form of the distribution of the service time.

We derive below an integral equation satisfied by the d.f. of a busy period, initiated by a single unit or customer.

Takacs Integral Equation

Let v be the service time,

$$B(t) = \Pr\{v \leqslant t\},\ B^*(s) = \int_0^\infty e^{-st}\, dB(t),$$

and T be the duration of the busy period,

$$G(t) = \Pr\{T \leqslant t\},\ G^*(s) = \int_0^\infty e^{-st}\, dG(t).$$

Let G^{n*} denote the n-fold convolution of G with itself. Let $A(x)$ denote the number of arrivals in $[0, x]$.

By conditioning on the duration of the service time v of the initial customer inaugurating the busy period and on the number A of arrivals during his service time, one gets,

$$G(t) = e^{-\lambda t} B(t) + \sum_{n=1}^{\infty} \int_0^t \Pr\{T \leqslant t \mid v = x, A(x) = n\}\, dB(x) \Pr\{A(x) = n\};$$

the first term corresponds to the case that none arrives during the service period of the initial customer (and then $x = t$). Now let T_i be the total service time of the ith descendant[†] and of all his progeny. Then T_i are i.i.d. and have the same distribution as T.

We have

$$\begin{aligned}\Pr\{T \leqslant t \mid v = x, A(x) = n\} &= \Pr\{T_1 + \ldots + T_n \leqslant t - x\}\\ &= G^{n*}(t - x),\ n \geqslant 1.\end{aligned}$$

Writing $G^{n*} = 1$, for $n = 0$, and noting that then $x = t$, we get,

$$G(t) = \sum_{n=0}^{\infty} \int_0^t \frac{e^{-\lambda x}(\lambda x)^n}{n!}\, G^{n*}(t - x)\, dB(x).$$

This equation is known as *Takacs integral equation*.

Putting the above expression for $G(t)$ on the r.h.s. of

$$G^*(s) = \int_0^{\infty} e^{-st}\, dG(t),$$

changing the order of integration in the repeated integral (on the r.h.s.) and on simplification, we get

$$G^*(s) = B^*(s + \lambda - \lambda\, G^*(s)),$$

which is a functional equation in terms of L.T. In view of its importance, we give an alternative derivation of the above equation.

Given that $v = x$, i.e. the initial customer ("ancestor") inaugurating the busy period departs at x, the number of ("descendants" i.e.) persons N arriving by time x has Poisson distribution with mean λx. Thus given $v = x$, the duration of the busy period T equals $x + S_N$, where

$$S_N = \sum_{i=1}^{N} T_i.$$

Since the L.T. of T is $G^*(s)$, the L.T. of S_N is

$$\exp\{-\lambda x + \lambda x\, G^*(s)\} \qquad \text{(by Theorem 1.11)}$$

Thus the L.T. of the r.v. T, given $v = x$, is

[†] In branching process terminology, the initial customer may be called "ancestor", those who arrive during his service time may be called (direct) "descendant" and the descendants of the descendants and so on, may be called "progeny"

$$E\{e^{-sT} \mid v = x\} = \exp(-sx)\exp\{-\lambda x + \lambda x\, G^*(s)\}$$
$$= \exp\{-x(s + \lambda - \lambda\, G^*(s))\}.$$

Finally,

$$G^*(s) = \text{L.T. of } T = E\{e^{-sT}\}$$
$$= \int_0^\infty E\{e^{-sT} \mid v = x\}\, dB(x)$$
$$= \int_0^\infty \exp\{-x(s + \lambda - \lambda\, G^*(s))\}\, dB(x)$$
$$= B^*(s + \lambda - \lambda\, G^*(s)).$$

As can be imagined, such an equation would occur in branching processes. The reduction of the busy period to branching processes is due first to Good, whose heuristic arguments leading to the equation were given formal content by Takacs. We state the following proposition:

Let $H(s)$ be the L.T. of a distribution with mean b and let $\lambda > 0$, then the equation

$$K(s) = H(s + \lambda - \lambda\, K(s))$$

has a unique root $K(s) \leqslant 1$; further, $K(s)$ is the L.T. of a proper probability distribution iff $\lambda b \leqslant 1$.

(For a proof of the above proposition, see Feller, Vol. II.)

From the above proposition it is clear that $G^*(s)$ is the L.T. of a proper probability distribution if $\rho = \lambda/\mu \leqslant 1$. If $\rho > 1$, there is a non-zero probability of a never-ending busy period, the probability being equal to the smallest positive root of the equation $\zeta = B^*(\lambda - \lambda\zeta)$.

Takacs obtains the inversion of $G^*(s)$ in the form

$$G(t) = \sum_{n=1}^{\infty} \int_0^t e^{-\lambda t} \frac{(\lambda t)^{n-1}}{n!}\, dB^{n*}(t).$$

Even without finding $G(t)$, one can find the moments of T from the functional equation in G^*. As can be easily verified

$$E(T) = \frac{E(v)}{1 - \rho}$$

and

$$\operatorname{var}(T) = \frac{E(v^2)}{(1-\rho)^3} - \frac{(E(v))^2}{(1-\rho)^2}.$$

As a special case, we can obtain $G^*(s)$ for $M/M/1$ model from the functional equation by putting $B^*(s) = \mu/(s + \mu)$. The equation then reduces to a quadratic in $G^*(s)$, whose root is seen to be β/ρ (as given in equation (3.23)).

Number Served During a Busy Period

Denote by $N(T)$, the number of customers served during a busy period initiated by a single customer ("ancestor") and by $P(z)$, its p.g.f. Denote by $N(T_j)$, the number served during the total service time T_j of the jth direct descendant and of all his progeny (i.e. $N(T_j)$ equals 1 plus the number of progeny of the jth direct descendant). Clearly

$$N(T) = 1 + \sum_j N(T_j)$$

($N(T)$ gives the total number of persons ever born in the family including the ancestor before the family becomes extinct).

Given that the initial customer departs at x, the number $A(x)$ of arrivals (direct "descendants") is a Poisson variate with mean λx. Denoting $P(z) = \xi$, we get

$$E\{z^{N(T_j)} \mid v = x\} = e^{-\lambda x+\lambda x\xi}.$$

Hence

$$E\{z^{N(T)} \mid v = x\} = ze^{-x(\lambda-\lambda\xi)}.$$

Thus we get

$$\xi = E\{z^{N(T)}\} = \int_0^\infty E\{z^{N(T)} \mid v = x\}\, dB(x)$$

$$= z\int_0^\infty e^{-x(\lambda-\lambda\xi)}\, dB(x),$$

i.e.

$$\xi = zB^*(\lambda - \lambda\xi).$$

Thus ξ satisfies the above functional equation. It can be shown that the equation has a unique root $\xi(z)$ such that $\xi(0) = 0$.

A fundamental result concerning the joint distribution of the busy period T and the number N served therein for a $M/G/1$ queue has been given by Prabhu (1960, 1965); (see Enns (1969), Scott *et al.* (1972) for related trivariate distributions).

The model $M/G/\infty$: transient solution

Suppose that the input is Poisson with intensity λ, the service time v is general with distribution function B having mean $1/\mu$ and that there is an infinite number of channels so that an arriving unit is immediately taken into service whatever be the number present in the system. Let $N(t)$ denote the number present in the system (or number of busy channels) at epoch t, and $p_n(t) = \Pr\{N(t) = n\}$. Assume that initially the system is empty, i.e. $p_0(0) = 1$, $p_n(0) = 0$, $n \neq 0$. If $X(t)$ denotes the number of arrivals by time t, then $\Pr\{X(t) = k\} = e^{-\lambda t}(\lambda t)^k/k!$. If $Y(t)$ denotes the number of departures by time t, then $X(t)=N(t)+Y(t)$. By conditioning on the number of arrivals $X(t)$, we get

$$p_n(t) = \Pr\{N(t) = n\} = \sum_{k=n}^{\infty} \Pr\{N(t) = n \mid X(t) = k\}\cdot\Pr\{X(t) = k\}.$$

Let $a \equiv a(t) = \Pr$ {an arbitrary unit arriving during $[0, t]$ is still in service at epoch t}. Then, conditioning on the epoch of arrival, we get

$$a = \int_0^t \Pr\{\text{service time of the arriving unit exceeds } t - x \mid \text{arrival occurs at epoch } x\} \cdot \Pr\{\text{arrival occurs at } x\}\, dx.$$

Now, arrival is in accordance with a Poisson process and so, given that an arrival occurs in $[0, t]$, the distribution of the arrival time is uniform in $[0, t]$ (Theorem 4.5). Thus,

$$a = \int_0^t [\Pr\{v > t - x\}\,(1/t)]\, dx$$

$$= (1/t) \int_0^t \{1 - B(t - x)\}\, dx = (1/t) \int_0^t \{1 - B(u)\}\, du$$

and
$$1 - a = (1/t) \int_0^t B(u)\, du.$$

Now, $\Pr\{N(t) = n \mid X(t) = k\} = \Pr$ {given that k arrived in $[0, t]$, the number in service at epoch t is n}

$$= \binom{k}{n} a^n (1 - a)^{k-n}, \qquad k \geqslant n.$$

Thus

$$p_n(t) = \sum_{k=n}^{\infty} \binom{k}{n} a^n (1 - a)^{k-n} e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

$$= \left[\frac{e^{-\lambda at} (\lambda at)^n}{n!}\right] \sum_{k=n}^{\infty} \frac{[\lambda t\,(1 - a)]^{k-n}}{(k - n)!}$$

$$= \frac{e^{-\lambda at} (\lambda at)^n}{n!}, \quad n \in N = \{0, 1, 2, \ldots\}$$

$$= \exp\left\{-\lambda \int_0^t \{1 - B(u)\}\, du\right\}\left\{\left[\lambda \int_0^t \{1 - B(u)\}\, du\right]^n \Big/ n!\right\}.$$

That is, the distribution of $N(t)$ is Poisson with mean

$$\lambda at = \lambda \int_0^t \{1 - B(u)\}\, du.$$

Similarly, it can be shown that the distribution of the departure process $Y(t)$ is also Poisson with mean

$$\lambda(1 - a)\, t = \lambda \int_0^t B(u)\, du.$$

Particular case: (1) *Steady state solution.*

As $t \to \infty$,
$$at = \int_0^t \{1 - B(u)\}\, du \to \frac{1}{\mu},$$

so that, $\lambda at \to \lambda/\mu = \rho$. Then the steady-state distribution of the number in the system in $M/G/\infty$ model is Poisson with mean $\rho = \lambda/\mu$, irrespective of the value of ρ.

(2) $M/M/\infty$ in transient state with $p_0(0) = 1$. Here $B(t) = 1 - e^{-\mu t}$ so that $\lambda at = (\lambda/\mu)\{1 - e^{-\mu t}\}$. Thus, for $n \in N$

$$p_n(t) = \left[\exp\left\{-\frac{\lambda}{\mu}(1 - e^{-\mu t})\right\}\right]. \left\{\frac{\lambda}{\mu}(1 - e^{-\mu t})\right\}^n \frac{1}{n!}.$$

As $t \to \infty$, $p_n(t) \to e^{-\rho}\rho^n/n!$, $n \in N$.

Thus, the steady-state distribution of the number in the system in a model with Poisson input is, irrespective of the form of the distribution of the service time as well as of the value of ρ, always Poisson in case of infinite channel model. As mentioned in Note (1) to Sec. 10.4 D, the distribution is always truncated Poisson in case of a s-channel (loss) model.

An Application in Reliability: *A spares inventory model* (Barlow and Proschan, 1975).

Consider a maintenance depot with a large number of repairmen. Suppose that the failure of units occur in accordance with a Poisson process with rate λ Failed units are brought to the depot for repair, the repair time v being i i.d. random variables having common distribution function $B(t) = \Pr\{v \leqslant t\}$ and mean $1/\mu$. Suppose that repair times are also independent of the arrival times and that the repaired units become part of the spares inventory, which consists initially (at $t = 0$), of a certain number of spares. Let $N(t)$ denote the number of repairmen engaged at repair at epoch t and let $N(0) = 0$.

Then we have an $M/G/\infty$ model for the spares inventory and the distribution of $N(t)$ is Poisson with mean $\lambda \int_0^t \{1 - B(u)\}\, du$.

If n is the number of spares in the inventory at $t = 0$, then the expected number of customers waiting for spares at epoch t is given by

$$E\,[\max\{(N(t) - n), 0\}]$$

$$= \sum_{k=n+1} (k - n) \cdot \Pr\{N(t) = k\}.$$

As $t \to \infty$, this expected number tends to

$$\sum_{k=n+1}^{\infty} (k - n)\frac{e^{-\rho}\rho^k}{k!}.$$

B. The Model $GI/M/1$

Here we assume that the service time distribution is exponential with mean $1/\mu$ and that the interarrival time is a random variable u, having an arbitrary (general) distribution with mean $1/\lambda$. Denote the d.f. of u by $A(t)$ and its p.d.f., when it exists, by $a(t)$; its L.T. is given by

$$A^*(s) = \int_0^\infty e^{-st}\, dA(t) \tag{6.17}$$

so that $$A^{*(k)}(0) \equiv \frac{d^k}{ds^k} A^*(s)\Big|_{s=0} = (-1)^k E(u^k) \tag{6 18}$$

and for $k = 1$, $A^{*(1)}(0) = -1/\lambda$.

Let t_n, $n = 1, 2, \ldots$, $(t_0 = 0)$ be the epoch at which the nth arrival occurs (or nth unit arrives). The process $N(t_n - 0) = Y_n$, $n = 0, 1, 2, \ldots$ gives the number in the system immediately before the arrival of the nth unit. Then

$$Y_{n+1} = Y_n + 1 - B_{n+1}, \text{ if } Y_n \geqslant 0,\ B_{n+1} \leqslant Y_{n+1},$$

where B_{n+1} is the number of units served during $(t_{n+1} - t_n)$, i.e. the interarrival time between the nth and $(n+1)$st unit. As B_n is independent of n, i.e. $B_n = B$ for all n, its distribution is given by

$$g_r = \Pr\{B = r\} = \int_0^\infty \frac{e^{-\mu t}(\mu t)^r}{r!}\, dA(t),\ r = 0, 1, 2, \ldots \tag{6.19}$$

The arrival point conditional probabilities

$$p_{ij} = \Pr\{Y_{n+1} = j \mid Y_n = i\}$$

are given by
$$\begin{aligned} p_{ij} &= g_{i+1-j},\ i + 1 \geqslant j \geqslant 1,\ i \geqslant 0 \\ &= 0, \quad i + 1 < j, \end{aligned} \tag{6.20}$$

and
$$\sum_{j=0}^{i+1} p_{ij} = 1.$$

Thus
$$\begin{aligned} p_{i0} &= 1 - \sum_{j=1}^{i+1} p_{ij} = 1 - \sum_{j=1}^{i+1} g_{i+1-j} \\ &= 1 - \sum_{r=0}^{i} g_r = h_i \text{ (say)}, \quad i \geqslant 0. \end{aligned} \tag{6.21}$$

Since all p_{ij} depend only on i and j, $\{Y_n, n \geqslant 0\}$ is a Markov chain having t.p.m.

$$P = (p_{ij}) = \begin{pmatrix} h_0 & g_0 & 0 & 0 & 0 & \cdots \\ h_1 & g_1 & g_0 & 0 & 0 & \cdots \\ h_2 & g_2 & g_1 & g_0 & 0 & \cdots \\ \cdots & & \cdots & & & \cdots \\ \cdots & & \cdots & & & \cdots \end{pmatrix}. \tag{6.22}$$

As $g_r > 0$, the chain is irreducible and aperiodic. It can also be shown that it is ergodic, persistent non-null when $\rho < 1$. Thus, when $\rho < 1$, i.e. in the ergodic case, the limiting arrival point system size probabilities

$$v_j = \lim_{n\to\infty} p_{ij}^{(n)}$$

exist and are given as the unique solution of the system of equations

$$V = VP, \tag{6.23}$$

where $V = (v_0, v_1, \ldots), \Sigma\, v_j = 1.$

We now proceed to find V.

Let $G(s) = \sum_{r=0}^{\infty} g_r\, s^r$ be the p.g.f. of $\{g_r\}$, i.e. of the r.v. B, the number of units served during an interarrival interval. We have

$$\begin{aligned} G(s) &= \sum_{r=0}^{\infty} g_r s^r \\ &= \sum_{r=0}^{\infty} s^r \int_0^{\infty} \frac{e^{-\mu t}\,(\mu t)^r}{r!}\, dA(t) \\ &= \int_0^{\infty} e^{-\mu(1-s)t}\, dA(t) \\ &= A^*(\mu(1-s)). \end{aligned}$$

Also
$$\begin{aligned} E(B) &= G'(1) = -\mu A^{*(1)}(0) \\ &= \mu/\lambda = 1/\rho. \end{aligned}$$

The equations (6.23) can be written as

$$v_0 = \sum_{r=0}^{\infty} v_r\, h_r \tag{6.24}$$

and
$$v_j = \sum_{r=0}^{\infty} v_{r+j-1}\, g_r, \quad j \geqslant 1. \tag{6.25}$$

Denoting the displacement operator by E (so that $E^r(v_k) = v_{k+r}$ etc.), we can write (6.25) as a difference equation

$$E(v_{j-1}) = v_j = \sum_{r=0}^{\infty} g_r \cdot E^r(v_{j-1}), \; j \geqslant 1$$

or,
$$\{E - \sum_{r=0}^{\infty} g_r\, E^r\}\; (v_{j-1}) = 0$$

or,
$$\{E - G(E)\}\; v_{j-1} = 0, \; j-1 \geqslant 0. \tag{6.26}$$

This equation can be solved by the method given in Sec. 1.5. The characteristic equation of the difference equation is

$$r(z) = z - G(z) \equiv z - A^*(\mu - \mu z) = 0.$$

It can be shown that when $G'(1) = 1/\rho > 1$, i.e. $\rho < 1$, $r(z)$ has only one zero inside $|z| = 1$. Assume that $\rho < 1$, then, if the root of $r(z) = 0$

inside $|z| = 1$ is denoted by r_0 and the roots on and outside $|z| = 1$ are denoted by $s_0, s_1, s_2, \ldots$ then the solution of (6.26) is

$$v_j = C_0 r_0^j + \sum_i D_i s_i^j, j \geqslant 0,$$

where C_0, D_i are constants.

But since $\sum_j v_j = 1$, $D_i = 0$ for all i. Thus, when $\rho < 1$, we get,

$v_j = C_0 r_0^j, j \geqslant 0$ and from $\sum_{j=0}^{\infty} v_j = 1$, we get $C_0 = 1 - r_0$, so that

$$v_j = (1 - r_0) r_0^j, \quad j \geqslant 0. \tag{6.27}$$

The steady-state arrival point system size has a geometric distribution with mean $r_0/(1-r_0)$, r_0 being the unique root of $r(z) = 0$ lying inside $|z| = 1$.

Waiting time distribution: As the distribution of $\{v_n\}$ is geometric and the service time is exponential, the waiting time in the queue W_Q, being random geometric sum of exponential variables, is *modified* exponential. We have,

$$\Pr(W_Q = 0\} = 1 - r_0$$

and

$$\Pr\{W_Q \leqslant t\} = 1 - r_0 e^{-\mu(1-r_0)t}, t > 0. \tag{6.28}$$

The conditional distribution of the waiting time in the queue W_Q of those units that have to wait, is exponential with mean $1/\mu (1 - r_0)$.

Note: (1) The unique root r_0 $(0 < r_0 < 1)$ of $r(z) \equiv z - G(z) = 0$ can be obtained numerically by employing Newton-Raphson or other standard method.

(2) Kingman (1963) gives an explanation for the occurrence of geometric distribution as steady state solution of the $GI/M/1$ queue and discusses about the close formal connection between the theory of this queueing model and that of branching processes.

Example 6(c). *The model $E_k/M/1$.* Here the distribution of the inter-arrival interval is Erlang-k with mean $1/\lambda$ and L.T. $A^*(s) = \left(\frac{\lambda k}{s + \lambda k}\right)^k$

Thus, the characteristic equation becomes

$$z - G(z) \equiv z - A^*(\mu - \mu z) = z - \left(\frac{\lambda k}{\mu - \mu z + \lambda k}\right)^k = 0.$$

If the unique zero of $z - G(z)$ lying inside $|z| = 1$ is denoted by r_0, then $v_n = (1 - r_0) r_0^n, n \geqslant 0$.

Note: As in the case of $M/E_k/1$, the queueing process in $E_k/M/1$ could also be viewed as non-birth-death Markovian and the solution obtained from the corresponding C.K. equations.

Particular cases: (1) $M/M/1$: When $k = 1$,

$$z - K(z) = 0 \text{ gives } (z - 1)(z - \rho) = 0.$$

Thus when $\rho < 1$, the required root r_0 is ρ and we get

$$v_n = (1-\rho)\,\rho^n, \quad n \geqslant 0.$$

In this case $v_n = p_n$ for all n ($v_n = p_n$ only when $E_k \equiv M$, i.e. $k = 1$).

(2) $D/M/1$: When $k \to \infty$, $A^*(s) \to e^{-s/\lambda}$

and $$z - G(z) = 0 \text{ implies } ze^{(1-z)/\rho} = 1.$$

If the unique root inside $|z| = 1$ is denoted by ξ, then

$$u_n = (1-\xi)\,\xi^n, \quad n > 0.$$

Remarks: We take as regeneration points the arrival epochs in case of $GI/M/1$ model and the departure epochs in case of $M/G/1$ model. The queueing processes occurring in the two models $M/G/1$ and $GI/M/1$ are non-Markovian; however by extracting processes at these regeneration points, it has been possible to obtain (embedded) Markov chains from those processes. The embedded Markov chains indicate *system state at regeneration* points. What one needs also is the distribution of *general (or random) time system state.*

Let us now consider the queueing process of $GI/M/1$, having for its embedded Markov chain $\{Y_n\}$, where Y_n is the system size immediately prior to nth arrival. We have obtained $\{v_j\}$, the limit distribution of $\{Y_n\}$. Define $Z(t) = Y_n$, $t_n \leqslant t < t_{n+1}$. Then $Z(t)$, which gives the system size at the most recent arrival, is a semi-Markov process having $\{Y_n\}$ for its embedded Markov chain. $\{Y_n, t_n\}$ is an irreducible Markov renewal process.

Let $N(t)$ denote the system size at an arbitrary (general) time t. Denote

$$f_{ij}(t) = \Pr\{Z(t) = j \mid Z(0) = i\}$$
$$p_{ij}(t) = \Pr\{N(t) = j \mid N(0) = i\};$$

then $f_j = \lim\limits_{t\to\infty} f_{ij}(t)$ gives the limiting probability that the system size of the s—M.P. $Z(t)$ is j, whereas $p_j = \lim p_{ij}(t)$ as $t \to \infty$, gives the limiting probability that the system size of the general time process $N(t)$ is j.

We have to look for relationship, if any, existing between the three limit distributions $\{v_j\}$, $\{f_j\}$ and $\{p_j\}$.

From Theorem 7.1, we find that v_j and f_j are related as follows:

$$f_j = \frac{v_j M_j}{\sum\limits_i v_i M_i}, \tag{6.29}$$

where M_i is the the expected time spent in the state i during each visit.

The transitions, in case of this model, occur at the arrival points which are its regeneration points. Thus M_i, the expected time spent in a state i during each visit is the expected interarrival time. In other words, $M_i = 1/\lambda$ for all i. Putting the value of M_i in (6.29), we get

$$f_j = \frac{v_j/\lambda}{\sum\limits_i v_i/\lambda} = \frac{v_j}{\sum\limits_i v_i} = v_j \quad \text{for all } j,$$
$$= (1-r_0)\,r_0^j \quad \text{(from (6.27))}.$$

In order to obtain p_j in terms of f_j (and in terms of v_j) we have to make use of the relationship (obtained by Fabens, 1961) between f_j and p_j.

It is shown that for this particular model,

$$p_j = (\lambda/\mu)\, f_j/r_0,$$
$$= (\lambda/\mu)\,(1 - r_0)\, r_0^{j-1}, \qquad j \geqslant 1$$

and
$$p_0 = 1 - \lambda/\mu.$$

C. The Model $M/G^{a,b}/1$

Model with Poisson input, general service time and under general bulk service rule. As in Sec. 10.6 A, we easily see that the departure point process $\{X_n = N(t_n + 0)\}$ is a Markov chain.

Denote $q_r \equiv \Pr\{\text{Number of arrivals during a service time} = r\}$

$$= \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^r}{r!}\, dB(t).$$

We have $Q(s) = \sum_r q_r s^r = B^*(\lambda - \lambda s)$, where $B^*(s) = \int_0^\infty e^{-st}\, dB(t)$. The transition probabilities $p_{ij} = \Pr\{X_{n+1} = j \mid X_n = i\}$ of the chain $\{X_n, n \geqslant 0\}$ are given by

$$\begin{aligned} p_{ij} &= q_{j-i+b}, & i > b,\ j \geqslant i - b \\ &= 0, & i \geqslant b,\ j < i - b \\ &= q_j, & a \leqslant i \leqslant b,\ j \geqslant 0 \end{aligned}$$

and
$$p_{ij} = q_j \qquad , 0 \leqslant i \leqslant a - 1,\ j \geqslant 0.$$

It is clear that the p_{ij}'s are independent of a; they are also independent of i for $0 \leqslant i \leqslant b$ and thus have the same values for all $i = 0, 1, \ldots, b$. The first $(b + 1)$ rows of the t.p.m are identical. The t.p.m. $P = (p_{ij})$ is given by

$$\begin{array}{c|cccc} & 0 & 1 & 2 & 3\ \ldots \\ \hline 0 & q_0 & q_1 & q_2 & q_3 \cdots \\ 1 & q_0 & q_1 & q_2 & q_3 \cdots \\ 2 & q_0 & q_1 & q_2 & q_3 \cdots \\ \vdots & & & \cdots & \\ b & q_0 & q_1 & q_2 & q_3 \cdots \\ b+1 & 0 & q_0 & q_1 & q_2 \cdots \\ b+2 & 0 & 0 & q_0 & q_1 \cdots \\ \vdots & \cdots & & \cdots & \cdots \\ & \cdots & & \cdots & \cdots \end{array}$$

The chain is irreducible and aperiodic. When the traffic intensity ρ (defined by $\rho = \lambda/b\mu$) is less than unity, the chain is ergodic. Then the limiting or steady-state departure point probabilities

$$v_j = \lim_{n \to \infty} p_{ij}^{(n)}, \ j = 0, 1, 2, \ldots$$

exist and $V = (v_0, v_1, \ldots)$ is given as the unique solution of $V = VP$. Thus, we have, for $j = 0, 1, 2, \ldots$

$$v_j = (v_0 + v_1 + \ldots + v_{b-1})\, q_j + v_b q_j + v_{b+1}\, q_{j-1} + v_{b+2}\, q_{j-2} + \ldots + v_{b+j}\, q_0$$

$$= \Big(\sum_{r=0}^{b-1} v_r\Big)\, q_j + \Big(\sum_{r=0}^{j} q_{j-r}\, v_{b+r}\Big).$$

The p.g.f. of $\{v_j\}$ is given by

$$V(s) = \sum_{j=0}^{\infty} v_j s^j$$

$$= \Big(\sum_{r=0}^{b-1} v_r\Big)\Big(\sum_{j=0}^{\infty} q_j\, s^j\Big) + \sum_{j=0}^{\infty} \Big(\sum_{r=0}^{j} q_{j-r}\, s^{j-r}\, v_{b+r}\, s^r\Big)\}.$$

The first term on the r.h.s. is $\big(\sum_{r=0}^{b-1} v_r\big)\, Q(s)$; the second term is

$$\sum_j \{(q_j\, s^j)(v_b) + (q_{j-1}\, s^{j-1})(s v_{b+1}) + \ldots + (q_0)(s^j\, v_{b+j})\}$$

$$= Q(s)\, \{v_b + s v_{b+1} + \ldots\}$$

$$= Q(s) \left\{\frac{1}{s^b} \sum_{r=b}^{\infty} v_r\, s^r\right\}$$

$$= Q(s)\, (1/s^b) \left\{ V(s) - \sum_{r=0}^{b-1} v_r s^r \right\}.$$

Thus, $$(1 - Q(s)/s^b)\, V(s) = (Q(s)/s^b)\Big\{\sum_{r=0}^{b-1} (s^b - s^r)\, v_r\Big\}$$

or, $$V(s) = \frac{\sum_{r=0}^{b-1} (s^b - s^r)\, v_r}{s^b - Q(s)} \cdot Q(s)$$

$$= \frac{\sum_{r=1}^{b-1} (s^b - s^r)\, v_r}{\{s^b/Q(s)\} - 1}. \qquad (6.30)$$

Putting $B^*(\lambda - \lambda s)$ for $Q(s)$, we also get the following expression for $V(s)$.

$$V(s) = \frac{\sum_{r=0}^{b-1} (s^b - s^r)\, v_r}{s^b - B^*(\lambda - \lambda s)} \cdot B^*(\lambda - \lambda s) \qquad (6.31)$$

$$= \frac{\sum_{r=0}^{b-1} (s^b - s^r)\, v_r}{\{s^b/B^*(\lambda - \lambda s)\} - 1}. \qquad (6.31a)$$

Particular cases: (1) $M/G/1$. Let $b = 1$, then $a = 1$. Noting that $v_0 = 1 - \rho$, and putting $a = b = 1$, we find that the result (6.7) at once follows from (6.30).

(2) $M/M^{a,b}/1$: We have $Q(z) = B^*(\lambda - \lambda z) = \frac{\mu}{\lambda - \lambda z + \mu}$. The distribution $\{v_k\}$ of the *system size at departure instants* in steady state for $a = 0, 1, \ldots, b$ has the p.g.f.

$$\begin{aligned} V(z) &= \sum_{k=0}^{\infty} v_k z^k \\ &= \frac{\sum_{k=0}^{b-1} (z^b - z^k)\, v_k}{z^b\{1 + (\lambda/\mu)(1 - z)\} - 1}. \end{aligned} \tag{6.32}$$

Now the denominator of $V(z)$ has $(b-1)$ simple zeros $z = R_1, \ldots, R_{b-1}$ inside the unit circle $|z| = 1$. The b factors $(z - 1)\prod_{i=1}^{b-1}(z - R_i)$, common to both the numerator and the denominator cancel out, leaving out in the denominator a factor $(z - R_0)$, R_0 being the zero *outside* $|z| = 1$ of the function $z^b\{1 + (\lambda/\mu)(1 - z)\} - 1$ occurring in the denominator.

Thus, $\quad V(z) = \frac{C}{z - R_0}$ and $V(1) = 1$ gives $C = 1 - R_0$

so that $\quad V(z) = \frac{1 - R_0}{z - R_0} = \frac{(1 - r_0)}{1 - r_0 z}$,

where $r_0 = 1/R_0$ is the unique root inside $|z| = 1$ of the equation whose roots are the reciprocals of those of $z^b\{1 + (\lambda/\mu)(1 - z)\} - 1 = 0$, i.e. r_0 is the unique root *inside* $|z| = 1$ of $\lambda z^{b+1} - (\lambda + \mu)\, z + \mu = 0$.

Thus, for all $a = 0, 1, \ldots, b$, we get, from $V(z) = (1 - r_0)(1 - r_0\, z)^{-1}$,

$$v_k = (1 - r_0)\, r_0^{\,k}, \; k = 0, 1, 2, \ldots.$$

For example, for $b = 2$, $a = 0, 1, 2$ we get $r_0 = (2\lambda/\mu)/(1 + \sqrt{(1 + 4\lambda/\mu)}$ and $v_k = (1 - r_0)\, r_0^{\,k}$; for $b = 1$, $a = 0, 1$ we get $r_0 = \rho$, so that $v_k = (1 - \rho)\, \rho^k$.

Remarks: (1) From (6.30), it is clear that the p.g.f. $V(s)$ which is a function of b, is independent of a. Thus given b, one gets the same steady state departure point distribution of system size under the general bulk service rule for all values of $a\,(= 1, 2, \ldots, b)$, and also for $a = 0$, (i.e., intermittently available server). It is clear that though for different values of a, the actual processes differ, all the embedded Markov chains considered at the regeneration points (which are departure instants) are identical.

(2) To obtain the steady-state distribution $\{\pi_j\}$ of the general time system size $N(t)$, one can proceed as follows:

First, consider the embedded Markov chain $\{X_n\}$ at departure points, i.e. $X_n = N(t_n + 0)$ and obtain its steady state distribution $\{v_j\}$ (this has already been considered in the example).

Secondly, consider the embedded Markov chain $\{Y_n\}$ at not only its instants of departure, but also at instants when $N(t)$ reaches the level a

from below and the service commences. This will imply that transitions of Y_n will occur at all instants of departure as well as *all* instants of commencement of service (instants of departure and commencement of service will coincide when $N(t_i + 0) \geqslant a$). Deduce the steady-state distribution $\{r_j\}$ of $\{Y_n\}$ from that of $\{v_j\}$.

Thirdly, construct a semi-Markov process $Z(t)$, having $\{Y_n\}$ as its embedded Markov chain. Obtain the steady-state distribution $\{f_j)$ of $Z(t)$ in terms of $\{r_j\}$ using Theorem 7.2.

Finally, obtain the steady-state distribution $\{\pi_j\}$ of $N(t)$ by conditioning $N(t)$ with $Z(t)$, (with the help of a theorem of Fabens (1961)—based on an elegant probabilistic argument concerning the number of units arriving between the instant of the last transition before t and the instant t itself).

Note that the distribution $\{r_j\}$ and therefore $\{\pi_j\}$ will also depend on a, although that of $\{v_j\}$ *does* not.

See Fabens (1961) and for more details, Neuts (1967), Teghem *et al.* (1969). For an alternative approach using the technique of inclusion of supplementary variables, see Borthakur and Medhi (1974); see also Chaudhry and Templeton (1981), who obtain the p.g.f.'s of the queue size at random epochs and at departure epochs and a relationship between the two p.g.f.'s (for the $M/G^{1,b}/1$ model).

Some General Observations

The pace at which research in queueing theory has progressed amply demonstrates its fertility. It is clear that the theory is moving forward and that its applicability is being enhanced. The extensive literature has mainly been concerned with describing the stochastic processes of interest in various queueing models.

With a view to simplify the existing theory, various methods have been advanced. For example, random walk approach has been used to provide simple proof of some well-known results as well as to obtain new results on a variety of queueing problems. Reference may also be made to Wiener-Hopf techniques in queueing theory (Smith and Spitzer). While examining extensions of these techniques to continuous time, ladder processes have been investigated (Prabhu, 1970), Prabhu *et al.* (1971, 1973), and it is shown (Prabhu, 1974) how the techniques could be employed to simplify existing theory and to generate new results. Mention may also be made of the general definition and properties of the Markov renewal branching processes which have been shown to be useful in the study several queueing models (Neuts, 1974).

One topic of considerable mathematical and practical interest is the heavy traffic research, whose principal objectives are (i) to describe unstable queueing systems (having traffic intensity greater than unity), and (ii) to approximate highly saturated stable queueing systems (having traffic intensity less than but close to unity). The term heavy traffic is due to Kingman who initiated work in this topic (for an extensive recent bibliography, see Whitt (1974). The central theme for heavy traffic is limit theorems. Further areas of research are concerned with the rates of convergence, approximations for multi-channel queues, finite queues, control etc.

As for statistical analysis in queueing theory, pioneering work was done by Clarke, who began by obtaining the maximum likelihood estimators for the arrival and service parameters of the simple queue. Important contribution on estimation of parameters and confidence statements about them have been made by Cox (1965), Wolff (1965), Lilliefors (1966). Wolff extended the work of Billingsley on likelihood estimation of Markov chains in continuous time, including limit theorems and hypothesis testing and made use of the same to obtain results for birth and death queueing models.

The research in control and optimal operation of queueing systems is centred around optimal design and control of a model with a cost structure (for a review, see Stidham and Prabhu (1974), Sobel (1974)); for a classified bibliography with over 150 titles, see Crabill *et al.* (1977). The decision variables considered are arrival and service parameters, queue parameters, operating time parameters; the objective functions considered are concerned with average or discounted costs. The theory of Markov and semi-Markov decision processes have been used in various studies. Martingale approach has also been made in the formulation and solution of control problem in queues (Martins-Neto and Wong (1976)). The current trend is of piece meal and adhoc applications of general models. Stidham and Prabhu emphasise on the need for common methodology and indicate some direction in which a unified theory could be developed.

10.7 Reliability

A. Introduction

Investigation in reliability began in the 1950's, and the growth of the theory began to gain momentum after a decade or so (some standard works and survey papers are listed in the References).

Reliability theory is mainly concerned with the determination of the probability that a system, consisting possibly of several components, will operate adequately for a given period of time in its intended application.

Suppose that a component i of a system is capable of just two modes of performance—functioning and not functioning, which may be denoted by 1 and 0 respectively. The mode of performance of the component i may be represented by a Bernoulli random variable X_i, with values 1 and 0 depending on whether the component is functioning or not functioning. Suppose that

$$p_i = \Pr\{X_i = 1\} = 1 - \Pr\{X_i = 0\};$$

then p_i is the reliability of the ith component.

Suppose that a system has some finite number n, of independent components labelled $1, 2, \ldots, n$, and that the system is capable of just two modes of performances. Represent the mode of performance of the system by the Bernoulli r.v X. Suppose that, given the structure of a system, the knowledge of its performance can be determined from that of its components. The system structures generally considered, are described below.

(i) *Series system*: The system functions iff all the n components function. We have $X = \min(X_1, \ldots, X_n)$; the reliability of n components is given by

$$\begin{aligned} P = \Pr\{X = 1\} &= \Pr\{\min(X_1, \ldots, X_n) = 1\} \\ &= \Pr\{X_1 = 1, X_2 = 1, \ldots, X_n = 1\} \\ &= \prod_{i=1}^{n} p_i. \end{aligned} \tag{7.1}$$

(ii) *Parallel system*: The system functions iff at least one of the n components functions. We have $X = \max(X_1, \ldots, X_n)$. The system reliability is given by

$$\begin{aligned} P = \Pr\{X = 1\} &= \Pr\{\max(X_1, \ldots, X_n) = 1\} \\ &= 1 - \Pr\{X_1 = 0, X_2 = 0, \ldots, X_n = 0\} \\ &= 1 - \prod_{i=1}^{n} (1 - p_i). \end{aligned} \tag{7.2}$$

(iii) *k-out of n-system*: The system functions iff at least k $(1 \leqslant k \leqslant n)$ of the n components function. As particular cases, we get the series system for $k = n$ and the parallel system for $k = 1$. Suppose, for simplicity, that $p_i = p$ for all i, then X is a binomial variate with parameters n, p. Thus

$$\begin{aligned} P &= \Pr\{X = 1\} \\ &= \Pr\{X \geqslant k\} = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}. \end{aligned} \tag{7.3}$$

Lifetime distributions: Let us now consider time. Assume that a device functions for some time after which it fails, and then remains in that non-functioning state. The duration of time T, for which it functions is its lifetime. A number of failure distributions have been suggested as models to describe the lifetime of a device.

Definition : The *reliability* (or the *survival function*) of the device at time t is defined by

$$\begin{aligned} R(t) = \Pr\{T \geqslant t\} &= 1 - F(t), \quad t \geqslant 0 \\ &= 1 \qquad\qquad\quad t < 0, \end{aligned} \tag{7.4}$$

where F is the distribution function of T.

Denote by $f(t)$. the p.d.f. of T, when it exists.

Exponential distribution. which has the memoryless property, characterises the phenomenon of no wear. Future lifetime of a device remains the same irrespective of its previous use, iff, its lifetime distribution is exponential. But the phenomenon of wearout has also to be taken into account in certain cases. In an attempt to find out distributions which describe diverse phenomena associated with a device, certain characteristics have been evolved. We define below two of them.

Definition: *Hazard rate* (or failure rate) function of the distribution F is defined by the ratio

$$r(t) = \frac{f(t)}{1-F(t)} = \frac{f(t)}{\Pr\{T > t\}} = \frac{-R'(t)}{R(t)}, \tag{7.5}$$

for values of t for which $F(t) < 1$. Here $r(t)\,dt$ gives the probability that the device has a lifetime between t and $t + dt$, given that it has functioned upto time t. We have

$$R(t) = R(0) \exp\left\{-\int_0^t r(x)\,dx\right\}.$$

F is an increasing (decreasing) hazard rate IHR (DHR) distribution if $r(t)$ is an increasing (decreasing) function of $t \geqslant 0$. The hazard rate corresponds to age-specific death rate in demographic studies.

Definition: *Hazard rate average* is defined by

$$\frac{1}{t}\int_0^t r(x)\,dx.$$

F is an increasing hazard rate average (IHRA) function if the hazard rate average increases in $t \geqslant 0$. (A detailed account of IHRA distributions is given by Birnbaum, Esary and Marshall (1966).) We list below some of the distributions generally considered in the literature:

Distribution	$f(t)\,(t > 0)$	$r(t)$
Exponential	$ae^{-at}\,(a > 0)$	a
Weibull	$a\alpha t^{\alpha-1} e^{-at^{\alpha}}\,(a > 0, \alpha > 0)$	$a\alpha t^{\alpha-1}$
Gamma	$\dfrac{a(at)^{\alpha-1} e^{-at}}{\Gamma(\alpha)}$ $(a > 0, \alpha > 0)$	
Modified extreme value	$\dfrac{1}{a} \exp\left\{-\dfrac{(e^t - 1)}{a} + t\right\}$ $a > 0$	$\dfrac{e^t}{a}$

The exponential distribution is the only distribution having constant hazard rate, as is expected. The modified extreme value distribution has an IHR and Weibull distribution has IHR for $\alpha > 1$ and DHR for $0 < \alpha < 1$.

B. System Reliability

Suppose that a system consists of n independent components $1, 2, \dots i, \dots n$, and that the lifetime of the system is T, while that of ith component is T_i. The system reliability (or survival function) at time t is $R(t) = \Pr(T > t)$, and the reliability of ith component at time t is $R_i(t) = \Pr(T_i > t)$.

Expected lifetime is given by

$$E(T) = \int_0^\infty t\left(-\frac{dR}{dt}\right)dt$$

$$= -\int_0^\infty t\,dR(t) = \int_0^\infty R(t)\,dt.$$

(i) *Series system*: We have $T = \min(T_1, \dots, T_n)$.

Thus, $R(t) = \Pr\{T > t\} = \Pr\{\min(T_1, \dots, T_n) > t\}$

$$= \prod_{i=1}^{n} \Pr\{T_i > t\} = \prod_{i=1}^{n} R_i(t). \tag{7.6}$$

The hazard rate function of the system is

$$r(t) = -\frac{R'(t)}{R(t)} = \frac{\sum_{i=1}^{n}(-R_i'(t))\prod_{j\neq i}^{n} R_j(t)}{\prod_{j=1}^{n} R_j(t)}$$

$$= \sum_{i=1}^{n} \frac{-R_i'(t)}{R_i(t)} = \sum_{i=1}^{n} r_i(t),$$

where $r_i(t)$ is the hazard rate function of the component i. Thus, if the lifetime distribution of each component is an IFR, then that of the series system is also so.

(ii) *Parallel system*: We have $T = \max(T_1, \dots, T_n)$.

Thus, $R(t) = \Pr\{T > t\} = \Pr\{\max(T_1, \dots, T_n) > t\}$

$$= 1 - \prod_{i=1}^{n} \Pr\{T_i < t\} = 1 - \prod_{i=1}^{n}(1 - R_i(t)). \tag{7.7}$$

For parallel system, the lifetime distribution of the system is not necessarily IFR, even if that of each component is.

(iii) *k-out of n-system*: The system lifetime, is the corresponding order statistics

$$T^{(1)}, T^{(2)}, \dots, T^{(k)}, \dots, T^{(n)},$$

(formed from $T_1, \dots, T_n$) for $k = 1, 2, \dots, n$.

Example 7(a). Consider a system consisting of n components such that the failure of the ith component occurs in accordance with a Poisson process of intensity a_i. To find the reliability of the system under different system structures. Here T_i is exponential with mean $1/a_i$, i.e. $R_i(t) = e^{-a_i t}$.

(i) *Series system*

$$R(t) = \Pi\, R_i(t) = e^{-(\Sigma a_i)t}.$$

T is exponential with mean $1/\Sigma a_i$.

(ii) *Parallel system*

$$R(t) = 1 - \prod_{i=1}^{n} (1 - e^{a_i t})$$

$$= \Sigma\, e^{a_i t} - \Sigma\, e^{-(a_i + a_j)t} + \ldots \pm e^{-(\Sigma a_i)t},$$

and
$$E(T) = \Sigma \frac{1}{a_i} - \Sigma \frac{1}{(a_i + a_j)} + \ldots \pm \frac{1}{\Sigma a_i}.$$

In particular, if $a_i = a$ for all i, then

$$R(t) = 1 - (1 - e^{-at})^n = \sum_{r=1}^{n} (-1)^{r-1} \binom{n}{r} e^{-rat};$$

also
$$R(t) = \sum_{r=1}^{n} \binom{n}{r} (e^{-at})^r (1 - e^{-at})^{n-r},$$

and
$$E(T) = \sum_{r=1}^{n} (-1)^{r-1} \binom{n}{r} \Big/ (ar).$$

By integrating in (0, 1), the identity

$$\sum_{r=0}^{n-1} (1 - t)^r = \{1 - (1 - t)^n\}/t,$$

we find that

$$E(T) = \frac{1}{a} \sum_{r=1}^{n} \frac{1}{r}.$$

For large n, $E(T) \to \frac{1}{a}(\log n + \gamma)$, where $\gamma = 0.577\ldots$, is Euler's constant.

(iii) *k-out of n-system.* Suppose that $a_i = a$ for all i. We have

$$R(t) = \Pr\{T > t\}$$

$$= \sum_{r=k}^{n} \binom{n}{r} (e^{-at})^r (1 - e^{-at})^{n-r}$$

and

$$E(T) = \int_0^\infty R(t)\, dt = \sum_{r=k}^{n} \binom{n}{r} \int_0^\infty e^{-atr} (1 - e^{-at})^{n-r}\, dt$$

$$= \sum_{r=k}^{n} \binom{n}{r} \int_0^1 y^{r-1} (1 - y)^{n-r} \frac{dy}{a} \quad \text{(putting } y = e^{-at}\text{)}$$

$$= \frac{1}{a} \sum_{r=k}^{n} \binom{n}{r} \beta(r, n-r+1) = \frac{1}{a} \sum_{r=k}^{n} \frac{n!}{r!(n-r)!} \cdot \frac{\Gamma(r)\,\Gamma(n-r+1)}{\Gamma(n+1)}$$

$$= \frac{1}{a} \sum_{r=k}^{n} \frac{1}{r}.$$

Example 7(b). Suppose that a system consists of n components 1, 2, . . ., i, . . ., n; the first r of them, 1, 2, . . ., r are all of different kinds and the remaining $(n-r)$ are similar but are of another kind. The system functions iff each of the r different components functions and at least one of the remaining $(n-r)$ similar components function. Suppose that the failure of the ith components occur in accordance with independent Poisson processes, with parameter a_i for $i = 1, 2, \ldots, r$, and with parameter b for $i = r+1, \ldots, n$.

Then

$$\begin{aligned} R_i(t) &= e^{-a_i t}, \text{ for } i = 1, \ldots, r \\ &= e^{-bt}, \text{ for } i = r+1, \ldots, n \end{aligned}$$

so that

$$\begin{aligned} R(t) &= \left\{ \prod_{i=1}^{n} R_i(t) \right\} \left\{ 1 - \prod_{j=r+1}^{n} (1 - R_j(t)) \right\} \\ &= \prod_{i=1}^{n} e^{-a_i t} \{1 - (1 - e^{-bt})^{n-r}\}. \end{aligned}$$

Markovian Models in Reliability Theory

In Example 7(a), the failure process for individual components is assumed to be Markovian. The assumption that the failure process and the repair process, whenever repairs are considered, are Markovian is very often made; in such cases it is possible to write down at once the Chapman-Kolmogorov (C K.) equation of the Markov process of the resulting system. From the solutions of these equations, one can find the system reliability. The following example will illustrate the method.

Example 7(c). To find the reliability of a system consisting of n identical components connected in parallel, such that the failure of a component occurs independently of the others and in accordance with a Poisson process with intensity a. (This reduces to Example 7(a) (ii) with $a_i = a$.)

The operating process of the system may be represented by a Markov process $\{X(t), t \geqslant 0\}$ with $(n+1)$ states $0, 1, \ldots, i \ldots, n$, the state i denoting that there are i failed components at the instant t. The state n is an absorbing state. The transition to state n denotes its first passage; and transitions from any of the other states to some other states avoid the state n.

Let $p_i(t) = \Pr\{X(t) = i\}$ and the initial conditions be $p_0(0) = 1$, $p_i(0) = 0$, $i \neq 0$. The reliability $R(t)$ of the system at time t is given by

$$R(t) = p_0(t) + p_1(t) + \ldots + p_{n-1}(t) = 1 - p_n(t).$$

It is clear that the process $\{X(t), t \geqslant 0\}$ is a birth process with state dependent rates λ_i given by $\lambda_i = (n-i)\, a$, $i = 0, 1, \ldots, n$, n being an absorbing state.

The C.K. equations of the system are thus given by

$$p_0'(t) = -\,na\,p_0(t)$$
$$p_i'(t) = -(n-i)\,a\,p_i(t) + (n-i+1)a\,p_{i-1}(t), \quad i = 1, 2, \ldots, n-1$$
$$p_n'(t) = ap_{n-1}(t).$$

The first equation can be solved at once and the other equations (which are linear) can then be solved recursively. Alternatively, the equations can be solved by considering their Laplace transforms. Denote the L.T. of $p_i(t)$ by $P_i(s)$. By taking the L.T. of the above equations and making use of the initial conditions, one gets

$$sP_0(s) = 1 - na\,P_0(s)$$
$$sP_i(s) = -(n-i)a\,P_i(s) + (n-i+1)a\,P_{i-1}(s), \quad i = 1, 2, \ldots, n-1$$
$$sP_n(s) = a\,P_{n-1}(s).$$

For the particular case, $n = 2$, we have

$$P_0(s) = \frac{1}{s+2a}$$

and

$$P_1(s) = \frac{2a}{s+a}\,P_0(s) = \frac{2a}{(s+a)\,(s+2a)} = 2\left(\frac{1}{s+a} - \frac{1}{s+2a}\right).$$

Taking the inverse L.T., we get

$$p_0(t) = e^{-2at}$$

and $\quad p_1(t) = 2(e^{-at} - e^{-2at});$

Hence $\quad R(t) = p_0(t) + p_1(t) = 2e^{-at} - e^{-2at}.$

In the general case, we get

$$P_0(s) = \frac{1}{s+na}$$

$$P_i(s) = \frac{(n-i+1)a}{s+(n-i)a}\,P_{i-1}(a), \quad i = 1, 2, \ldots, n-1$$

$$P_n(s) = \frac{a}{s}\,P_{n-1}(s).$$

Thus, $\quad P_n(s) = \frac{a}{s}\,P_{n-1}(s) = \frac{a}{s}\cdot\frac{2a}{s+a}\,P_{n-2}(s)$

$$\ldots \quad \ldots$$

$$= \frac{a(2a)\ldots(na)}{s(s+a)\ldots(s+na)}.$$

Resolving into partial fractions, we get

$$\frac{a(2a)\ldots(na)}{s(s+a)\ldots(s+na)} = \frac{A_0}{s} + \frac{A_1}{s+a} + \ldots + \frac{A_n}{s+na},$$

where $A_0, \dots$ are constants, which can be determined as follows. Multiplying both sides by s and letting $s \to 0$, we get $A_0 = 1$.

Multiplying both sides by $(s + ia)$ and letting $s \to -ia$, we get

$$A_i = \frac{a \cdot (2a) \dots (na)}{(-ia)(-(i-1)a) \dots (-a)(a)(2a) \dots (n-i)a}$$

$$= (-1)^i \frac{(n-i+1)(n-i+2) \dots n}{i!} = (-1)^i \binom{n}{i}.$$

Thus, $$P_n(s) = \frac{1}{s} - \frac{\binom{n}{1}}{s+a} + \dots + (-1)^i \frac{\binom{n}{i}}{s+ia} + \dots \pm \frac{\binom{n}{n}}{s+na}.$$

Inverting the L.T., we get

$$p_n(t) = 1 - \binom{n}{1} e^{-at} + \dots \pm \binom{n}{n} e^{-nat}$$

$$= (1 - e^{-at})^n.$$

Thus, $$R(t) = 1 - p_n(t) = 1 - (1 - e^{-at})^n.$$

Note : This result was obtained in Example 7(*a*) by the straightforward method which was much simpler. This alternative method is given only to illustrate the application of the theory of Markov process.

We now make the assumption that a failed component can be repaired and renewed. The system can then be described by a queueing model, and the methods of queueing theory can be applied in its study. To find the reliability $R(t)$ at time t, one needs the transient state probabilities. We consider an example below. For more general discussion refer to the standard texts on Reliability theory, such as Gnedenko *et al.*

Example 7(d). Parallel system with maintenance. Consider a two-component parallel system, with a single repairman available for maintenance. Suppose that the failure and repair processes are both Markovian with rates a and b respectively (see also Exercise 4.11).

Here the process $\{X(t), t \geqslant 0\}$, has 3 states $i = 0, 1, 2$, the state i denoting the number of failed units at the instant t. The state 2 is an absrobing state. This may be considered as a machine interference problem (see Sec 10.4); the process may be represented by a birth and death process having 3 states 0, 1, 2, the state 2 being an absorbing state. Let $p_i = \Pr\{X(t) = i\}$; the C.K. equations are given by

$$p_0'(t) = -2a\, p_0(t) + b\, p_1(t)$$

$$p_1'(t) = -(a+b)\, p_1(t) + 2a\, p_0(t)$$

$$p_2'(t) = a\, p_1(t).$$

Let the initial conditions be $p_0(0) = 1$, $p_i(0) = 0$, $i = 1, 2$.

The above equations can be easily solved by taking their Laplace transforms. Denote the L.T. of $p(t)$ by $P(s)$. Taking L. T. and using initial conditions, the first two equations can be written as

$$(s+2a)\,P_0(s) - b\,P_1(s) = 1$$

$$-2a\,P_0(s) + (a+b+s)\,P_1(s) = 0.$$

Solving, we get

$$P_1(s) = \frac{2a}{s^2+(3a+b)\,s+2a^2}\,,\quad P_0(s) = \frac{a+b+s}{s^2+(3a+b)\,s+2a^2}.$$

$P_1(s)$ can be written as

$$P_1(s) = \frac{2a}{(s-s_1)(s-s_2)}$$

where

$$\begin{matrix} s_1 \\ s_2 \end{matrix} = \frac{-(3a+b) \pm \sqrt{\{(a+b)^2+4ab\}}}{2}$$

Thus,

$$P_1(s) = \frac{2a}{s_1-s_2}\left(\frac{1}{s-s_1} - \frac{1}{s-s_2}\right)$$

and its inverse is

$$p_1(t) = \frac{2a}{s_1-s_2}\,(e^{s_1 t} - e^{s_2 t})$$

Again,

$$P_0(s) = \frac{a+b+s}{(s-s_1)(s-s_2)}$$

$$= \frac{1}{(s_1-s_2)}\left(\frac{a+b+s_1}{s-s_1} - \frac{a+b+s_2}{s-s_2}\right)$$

and its inverse is

$$p_0(t) = \frac{(a+b+s_1)\,e^{s_1 t} - (a+b+s_2)\,e^{s_2 t}}{s_1-s_2}.$$

The system reliability is given by

$$R(t) = p_0(t) + p_1(t)$$

$$= \frac{(3a+b+s_1)\,e^{s_1 t} - (3a+b+s_2)\,e^{s_2 t}}{(s_1-s_2)}.$$

Using the relation $s^2+(3a+b)\,s+2a^2 = (s-s_1)(s-s_2)$, $R(t)$ can also be expressed as

$$R(t) = \frac{s_1\,e^{s_2 t} - s_2\,e^{s_1 t}}{(s_1-s_2)}$$

C. Shock Models and Wear Processes

We consider here a model described by Cox, Esary, Epstein, Gaver and later exhaustively examined by Esary, Marshall and Proschan (1973) for the lifetime of a device. Here we give some of the interesting results obtained by Esary *et al.*, to which references may be made for proof of the theorems stated. Some of the results of Esary *et al.* have been generalised to cover the multivariate case in a recent paper by Marshall and Shaked (1979).

The device is subjected to shocks which occur in accordance with a

Poisson process with parameter a. Denote the probability that the device survives the first k shocks, $(k = 0, 1, 2, \ldots)$ by q_k. Clearly, the reliability of the device at time t is

$$\begin{aligned} R(t) = \Pr\{T > t\} &= \sum_{k=0}^{\infty} q_k\, e^{-at}(at)^k/k!, \quad t \geqslant 0 \\ &= 1 \qquad\qquad\qquad\qquad\quad , t < 0 \end{aligned} \tag{7.8}$$

Now, q_k being the probability of surviving k shocks, we have

$$1 \geqslant q_0 \geqslant q_1 \geqslant q_2, \ldots \tag{7.9}$$

and p_k, the probability of failure on the kth shock, is given by

$$p_k = q_{k-1} - q_k \quad , k = 1, 2, \ldots \tag{7.10}$$

and $$p_0 = 1 - \sum_{k=1}^{\infty} p_k = 1 - q_0.$$

The p.d.f. $f(t)$ of the lifetime T is given by

$$f(t) = -R'(t) = a \sum_{k=1}^{\infty} p_k\, e^{-at}(at)^{k-1}/(k-1)!, \; t > 0 \tag{7.11}$$

and $$\Pr\{T = 0\} = 1 - \lim_{\varepsilon\to 0} \int_{\varepsilon}^{\infty} f(t)\,dt = p_0, \quad t = 0.$$

Its hazard rate function $r(t)$ is given by

$$\begin{aligned} r(t) &= \frac{f(t)}{R(t)} \\ &= a\left[1 - \left\{\sum_{k=0}^{\infty} q_{k+1}(at)^k/k!\right\}\Big/\left\{\sum_{k=0}^{\infty} q_k\,(at)^k/k!\right\}\right], t > 0 \end{aligned}$$

so that $r(t) \leqslant a$.

The equality holds if $q_k = 0$ for all $k > 0$, i.e. if $q_0 = 1$ and then T has exponential distribution.

One can express q_k and p_k in terms of derivatives of $R(t)$ as follows: Denoting by $R^{(k)}(t)$, the kth derivative of $R(t)$, we have

$$q_k = \sum_{j=0}^{k} \binom{k}{j} R^{(j)}(0)/a^j,$$

$$p_k = -\sum_{j=1}^{k} \binom{k-1}{j-1} R^{(j)}(0)/a^j,$$

The jth moment μ_j, $j = 1, 2, \ldots$ of T is given by

$$\mu_j = \frac{j!}{a^j} \sum_{k=0}^{\infty} \binom{k+j-1}{k} q_k,$$

provided the series on the r.h.s. converges. Some of the properties of q_k, of interest in reliability, carry over to $R(t)$. One such property is as given below.

Theorem 10.1. $R(t)$ is IHR if q_k/q_{k-1} is decreasing in $k = 1, 2, \ldots$.

Models for q_k : Consider that damage is caused to the device by a shock. Suppose that the ith shock causes a random damage X_i and that damages are additive, so that the damage caused by the first k shocks is $X_1 + X_2 + \ldots + X_k$. Suppose that there is a capacity or threshold which the device can endure. The threshold may be considered to be a constant or to be a random variable ξ. We assume here that the capacity or threshold has a fixed value x. The device survives the kth shock if $X_1 + \ldots + X_k$ does not exceed the threshold value x, so that

$$q_k = \Pr\{X_1 + \ldots + X_k \leqslant x\}.$$

Now the nonnegative random variables X_i may be (i) i.i.d., (ii) independent but not identical (as successive shocks could cause increasing damage) and (iii) neither independent nor identical (as accumulation of damage may result in loss of resistance to further damage).

If X_i, $i = 1, 2, \ldots$ are independent but not identical (and have F_i for their d.f.), then q_k is given by their convolution

$$q_k = F_1 * \ldots * F_k.$$

If X_i are i.i.d with d.f. F, then the d.f. of $X_1 + \ldots + X_k$ is the kth convolution F^{k*} of F, $k = 1, 2, \ldots$. Then

$$q_k = F^{k*}(x), \ (x \text{ fixed}),\ k = 0, 1, 2, \ldots$$

and
$$R(t) = \sum_{k=0}^{\infty} e^{-at} \frac{(at)^k}{k!} F^{k*}(x).$$

The process represented by $R(t) \equiv R_x(t)$, which has x as a parameter, is a compound Poisson process.

The stochastic process $\{Z(t), t \geqslant 0\}$ representing the accumulated damage or wear in $(0, t)$ is called the *wear process*. For each t,

$$\Pr\{Z(t) \leqslant x\} = R_x(t) = R(t).$$

We state below an important theorem concerning $Z(t)$.

Theorem 10.2. Assume that $Z(t)$ satisfies the following conditions:
(i) $\{Z(t), t \geqslant 0\}$ is a Markov process,
(ii) $Z(0) = 0$, $Z(t+h) - Z(t) \geqslant 0$ for all $t, h \geqslant 0$, with probability one, and (iii) $\Pr\{Z(t+h) - Z(t) \leqslant u \mid Z(t) = u\}$ is decreasing both in Z and t for $t \geqslant 0$, $h \geqslant 0$, $u \geqslant 0$.
Then the first passage time $T_x = \inf\{t \colon Z(t) > x\}$ of the Markov process $Z(t)$, is IHRA.

Note: (1) The condition (ii) means that the device enters service at time 0 with no accumulated wear, and that wear is non-negative and the condition (iii) implies that, for fixed t, an accumulation to wear can only weaken a device and make it more prone to further wear and that given equal amounts of accumulated wear, a device older in age is more prone to additional wear than one younger in age. The three conditions seem to be fairly realistic.

(2) Compound Poisson processes are included amongst the Markov processes satisfying the conditions (ii) and (iii).

(3) When the threshold has a fixed value x, the first passage time T_x of the Markov proeess $Z(t)$ is equivalent to the lifetime T of the device.

Example 7(e). If

$$\begin{aligned} q_k &= 1,\ k \leqslant n \\ &= 0,\ k > n, \end{aligned}$$

then from (7.10) we get

$$\begin{aligned} p_k &= 1,\ k = n+1 \\ &= 0, \text{ otherwise}; \end{aligned}$$

from (7.11) we get

$$f(t) = \frac{ae^{-at}\,(at)^n}{n!},$$

so that T has a gamma distribution; in particular, if $n = 1$, i.e. $q_0 = 1$, $q_k = 0$, $k = 1, 2, \ldots$, then T has an exponential distribution.

Hammed and Proschan (1973, 75) have extended these shock models and wear processes to cover the cases in which shocks occur in accordance to non-homogeneous Poisson process and more general birth processes.

The discussions in this section would give an indication of the role of stochastic processes in reliability. There are various other models in reliability where interesting stochastic processes arise. Cinlar (1975) shows how the theory of Markov renewal processes could be applied in reliability.

Current Trends: Some surveys

The qualitative concepts of reliability are somewhat old but its quantitative aspects have been developed over the past two decades or so. The developments are quite varied; here we mention only a few recent survey papers of interest. All of them contain exhaustive and useful lists of uptodate references.

Barlow and Proschan (1976) indicate some directions of current lists academic research in reliability.

Reliability, which has long been considered as a measured of system effectiveness, is now considered as an incomplete measure. *Availability*, which is a combined measure of reliability and maintainability is now receiving wider attention and usage as a measure of maintained system of effectiveness. An elaborate survey of the literature with a classification of the problems is given by Lie *et al.* (1977).

Considerable attention has been paid to reliability optimisation. A special issue of IEEE Transactions, Vol. R-26 (1977) is devoted to this topic. Tillman *et al.* (1977) present a review of the theoretical optimisation techniques used in reliability, and provide an extensive bibliography.

EXERCISES

10.1 Consider the simple queue and let $N(t)$ denote the system size at time t. Find the transition probabilities of the Markov process $\{N(t), t \geqslant 0\}$:

$$p_{ij}(\Delta t) = \Pr\{N(t+\Delta t) = j \mid N(t) = i\}.$$

Write down the rate matrix aud hence derive the Kolmogorov equations.

10.2 Find the distribution of the number (Q) in the queue in steady state in a $M/M/1$ queue with rates λ, μ. Find $E(Q)$ and $\operatorname{var}(Q)$. Verify that $E(Q) = \lambda E(W_Q)$. Find also $\operatorname{var}(W_Q)$, and verify that $\operatorname{var}(W_Q) + \operatorname{var}(v) = \operatorname{var}(W_S)$, where v is the service time.

10.3 Obtain the distribution of the number in the system in steady state for $M/M/s$ model by considering it as a birth and death process.

10.4 Using Little's formula obtain the average waiting time in the system for $M/M/1/R$ model.

10.5 **Centralisation of counters**

Consider that two identical $M/M/1$ queueing systems with the same rates λ, μ (intensity $\rho = \lambda/\mu$) are in operation. Show that the steady state distribution of the number N in the two systems combined is given by

$$\Pr\{N = n\} = (n+1)(1-\rho)^2 \rho^n, \; n \geqslant 0.$$

Now suppose that it is decided to form a single queue instead of two queues but to retain the two servers. Show that the steady state distribution of the number M in the system is given by

$$\Pr\{M = n\} = \frac{2(1-\rho)\rho^n}{1+\rho}, \; n \geqslant 1$$

$$= \frac{1-\rho}{1+\rho}, \quad n = 0.$$

Compare some measures of effectiveness and performance of the two models.

10.6 **Output process of a Markovian queueing system**

Show that for a $M/M/1$ queueing system in steady state, the intervals between successive departures are i.i.d. random variables having the same distribution as the interarrival interval.

Show that the output distribution is identical to the input distribution also for the $M/M/s$ system in steady state. (In other words, the *output process* is identical to the *input Poisson process* for a $M/M/s$ $(s \geqslant 1)$ queueing system in steady state) (Burke, 1968).

10.7 Let

$$p(k, r) = \frac{e^{-r} r^k}{k!},$$

$$E(s, r) = \sum_{k=0}^{s-1} p(k, r).$$

(Tables of these functions are availables.)

Show that $E\{W_Q\}$ for an $M/M/s$ queue in steady state can be expressed as

$$E\{W_Q\} = \left[\frac{1}{s\mu(1-\rho)}\left\{1 + (1-\rho)\frac{E(s, r)}{p(s, r)}\right\}^{-1}\right],$$

$$\rho = \lambda/\mu s = r/s.$$

10.8 Shonick (1970) refers to some studies on hospital administration and reports the following conclusion generally arrived at in regard to occupancy of hospital beds:

When the number of beds is such that all the beds are rarely filled up (such a situation arises in emergency wards for non-scheduled and emergency admissions), then one may expect the distribution of the number of occupied beds to be Poisson ; on the other hand, when the beds are often all filled, Poisson does not give a good fit.

Assuming that demand for beds occurs in accordance with a Poisson process and that occupancy times are exponentially distributed, give a plausible explanation for the occurrence of the observed phenomena.

10.9 Combination of a loss and a delay (waiting) system (Shonick, 1970).

Consider that two types of patients arrive at a hospital having a fixed number of beds, say s. These are (i) elective customers (or physician generated) and (ii) emergency customers (nature generated), their arrivals being in accordance with Poisson processes with rates a_1 and a_2 per day respectively. Assume that the duration of hospitalisation of both types of patients has exponential distribution with mean $1/b$ day. Elective customers who donot find a bed vacant wait and form a queue, while emergency customers who donot find a vacant bed leave and are lost.

Show that, in steady state, (when $a_1 < bs$), the distribution of the number in the system N is given by

$$\Pr(N = n) = p_n = \frac{(a/b)^n}{n!} p_0, \qquad n = 1, 2, \ldots, s$$

$$= \frac{(a/b)^s (a_1/b)^{n-s}}{s^{n-s}(s!)} p_0, \qquad n = s+1, s+2, \ldots$$

where

$$p_0 = \left\{\sum_{r=0}^{s-1} \frac{(a/b)^r}{r!} + \frac{(a/b)^s}{s!}\left(1 - \frac{a_1}{bs}\right)\right\}^{-1}, \quad a = a_1 + a_2.$$

Find the expected number of patients (i) hospitalised and (ii) in the queue.

Obtain the corresponding results for $M/M/s/\infty$ and $M/M/s/s$ models as particular cases.

10.10 Consider a $M/G/s/s$-loss system in equilibrium. If the random variable N denotes the number of busy channels and $B(s) \equiv B(s, \lambda/\mu) = \Pr\{N = s\}$, then show that

$$E(N) = (\lambda/\mu)(1 - B(s)).$$

Suppose that x_i is the indicator variable of the ith randomly chosen channel; $x_i = 1$ or 0 according as the ith randomly chosen channel is busy or idle; and that $p_s(A)$ denotes the probability of the event A (with reference to a $M/G/s/s$-loss system in equilibrium). Then show that

$$p_s(x_1 = 1) = (\lambda/\mu)(1 - B(s))/s$$

$$\begin{aligned} p_s\{x_1 = 1, x_2 = 1, \ldots, x_k = 1\} \\ = B(s)/B(s-k), \quad 1 \leqslant k < s \end{aligned}$$

$$\begin{aligned} p_s\{x_{k+1} = 1 \mid x_1 = 1, x_2 = 1, \ldots, x_k = 1\} \\ = (\lambda/\mu)[1 - B(s-k)]/(s-k) \\ = p_{s-k}(x_1 = 1) \end{aligned}$$

(Kaufman, 1979).

10.11 Machine interference with m machines and s repairman. (a) Show that the mean arrival rate λ' is given by

$$\lambda' = \sum_{n=0}^{s} (m-n)\lambda p_n = \lambda(m - L)$$

and that
$$W = \frac{L}{\lambda'} = \frac{L}{\lambda(m-L)}.$$

(b) Denote by X, the expected number of machines in working order. Show that, when s is close to m, then

$$X \simeq \frac{m}{1 + \lambda/\mu}$$

and when s is small, then

$$X \simeq \frac{s}{(\lambda/\mu)}.$$

(c) Special case: $s = 1$. Show that the probability p_r that the number of machines in working order is r is given by

$$p_r = \frac{p(r, 1/\rho)}{1 - E(m+1, 1/\rho)}, \; 1 \leqslant r \leqslant m.$$

10.12 A mechanic looks after 8 automatic machines, A machine breaks down, independently of others, in accordance with a Poisson process, the average length of time for which a machine remains in working order being 12 hours. The duration of time required for repair of a machine has an exponential distribution with mean 1 hour. Find:

(a) the probability that 3 or more machines will remain out of order at the same time;

(b) the average number of machines in working order;

(c) for what fraction of time, on the average, the mechanic will be idle; and

(d) the average duration of time for which a machine is not in working order.

10.13 *Multistage Poisson queueing system with infinite number of servers*: Jackson *et al.* (1980) have considered a finite number k of $M/M/\infty$ queues *in tandem* as a mathematical model for patient progress in a certain disease. Let the arrival to the first stage be Poisson with mean rate λ, the service time at ith stage be exponential with mean $1/\mu_i$; let the probability that at time t there are n_i customers in stage i ($i = 1, 2, \ldots, k$) be $P(n_1, n_2, \ldots, n_k, t)$. Show that the differential equation of the system state is given by

$$\begin{aligned}\frac{\partial P(n_1, n_2, \ldots, n_k, t)}{\partial t} &= -\left(\lambda + \sum_{i=1}^{k} n_i \mu_i\right) P(n_1, n_2, \ldots, n_k, t) \\ &\quad + \delta(n_1)\,\lambda P(n_1 - 1, n_2, \ldots, n_k, t) \\ &\quad + \sum_{i=1}^{k-1} \delta(n_{i+1})(n_i + 1)\,\mu_i\, P(n_1, \ldots, n_i + 1, n_{i+1} - 1, \ldots, n_k, t) \\ &\quad + (n_k + 1)\,\mu_k\, P(n_1, n_2, \ldots, n_k + 1, t),\end{aligned}$$

where

$$\begin{aligned}\delta(n) &= 1, \quad n > 0 \\ &= 0, \quad n = 0.\end{aligned}$$

In particular, when $k = 2$, show that the equation admits of the solution

$$P(n_1, n_2, t) = \left\{\frac{e^{-a_1} a_1^{n_1}}{n_1!}\right\}\left\{\frac{e^{-a_2} a_2^{n_2}}{n_2!}\right\},$$

where

$$\begin{aligned}a_1 &= (\lambda/\mu)\,(1 - e^{-\mu t}) \\ a_2 &= (\lambda/\mu)\,(1 - e^{-\mu t} - \mu t\, e^{-\mu t}).\end{aligned}$$

Deduce the steady state results.

10.14 $M/M^{a,b}/1$: Busy period distribution.

A busy period starts at the instant the idle server takes a batch for service leaving an empty queue and ends at the instant the server completes service of a batch leaving, for the first time, r units ($0 \leqslant r \leqslant a - 1$) in the queue.

Denote by $f_r(t)$, the p.d.f. of the busy period ending with r units in the queue, $r = 0, 1, 2, \ldots, a - 1$.

Show that

$$f_r(t) = \sum_{k=0}^{\infty} \left[(\mu)^{k+1} \int_0^t \exp\{-(\mu t + \lambda u)\} \frac{(t-u)}{k!} \right.$$

$$\left. \left\{ \frac{ka+r}{p^{ka+r}} I^{a}_{ka+r} (2\lambda p u) - \frac{ka+r+1}{p^{ka+r+1}} I^{b}_{ka+r+1} (2\lambda\, pu) \right\} \right] du$$

where $\qquad p^{b+1} = \mu/\lambda,$

and $\qquad I^{m}_{s}(x) = \sum_{n=0}^{\infty} (x/2)^{s+n(m+1)} [1/n!\,(mn+s)],$

s being a positive integer.

Obtain the busy period distribution of the simple queue $M/M/1$ as a particular case (Neuts, 1967; Borthakur, 1975).

10.15 The $M/M^{a,b}/s$ queue (s-channel Markovian queue under the general bulk service rule).

Let $P_{r,n}$ denote the steady state probability that r channels are busy and n units are waiting in the queue. Find the difference-equations of the system in steady state. Show that

$$P_{p,n} = A r_0^n, \quad n \geqslant 0,$$

where A is a constant and r_0 is the unique root in (0, 1) of the equation

$$\lambda - (\lambda + s\mu)\, x + s\, \mu x^{b+1} = 0.$$

10.16 The system $M/G^{a,a}/1$ (Single server system with Poisson input with service in batches of fixed size a). Let X_n, Y_n, $Z(t)$ denote the same processes as indicated in Sec. 10.6(c).

The Markov chain $\{X_n\}$ has the steady state distribution $\{v_k\}$ having p.g.f. as given by (6.31) with $b = a$.

Show that the Markov chain $\{Y_n\}$ has steady state distribution $\{r_k\}$,

where

$$\begin{aligned} r_k &= c v_k, & i &< a \\ &= c(v + v_k), & k &= a \\ &= c v_k, & k &> a \end{aligned}$$

and $\qquad v = \sum_{j=0}^{a-1} v_j, \qquad c = 1/(1+v).$

The semi-Markov process $\{Z(t)\}$ has steady state distribution

$$f_j = \lim_{t\to\infty} \Pr\{Z(t) = j\} = \mu_j r_j / \Sigma \mu_i r_i,$$

μ_j being the mean waiting time in the state j (Fabens, 1961).

10.17 $M^{(X)}/G^{a,b}/1$ model: Suppose that customers arrive in batches of size $r \geqslant 1$, the probability that a batch of r arrives in an infinitesimal interval of length dt being $\lambda c_r dt$ $(\sum_r c_r = 1)$. Suppose that service is in groups of size $k (a \leqslant k \leqslant b)$, and that $\eta(x)\, dx$ is the conditional probability that the service will be completed in $(x, x+dx)$, given that it has not been completed till time x. Let

$V_r(t)$ = Probability that at the instant 't', the queue length is equal to r and the service channel is idle (it is non-zero only when $0 \leqslant r \leqslant a-1$);

$U_n(t)$ = Probability that at the instant 't', the queue length is equal to n and the service channel is busy $(n \geqslant 0)$;

$U_n(x, t)\,dx$ = Probability that at the instant 't', the queue length is equal to 'n' and a batch is being serviced with elapsed service time lying between x and $x+dx$ $(n \geqslant 0)$, so that

$$U_n(t) = \int_0^\infty U_n(x, t)\,dx.$$

Obtain the differential-difference equations governing the system in terms of the above. Deduce the steady state relations.

10.18 Average cost in a batch service queue:

Suppose that units arrive at a service facility according to a Poisson process and that the single server is an infinite capacity batch server. When the server is free to serve, and there are fewer than q $(\geqslant 1)$ units waiting, he does not start service; but if there are q or more waiting, the service of all those waiting is initiated in a batch.

Let λ = rate of arrival,

v = service time of a batch,

k = fixed cost of each service,

h = waiting cost of a unit per unit time,

X_n = length of time between the start of nth and $(n+1)$st batch services, and

Y_n = cost incurred during nth service.

Show that $\{X_n, Y_n\}$ is a renewal reward process and that the expected long run expected cost $\zeta(q)$ per unit time is given by

$$\zeta(q) = \frac{E(Y)}{E(X)} = \frac{k + \lambda h E(v^2) + (h/2\lambda) \sum_{n=0}^{q-1} p_n (q^2 - q - n^2 + n)}{E(v) + \sum_{n=0}^{q-1} p_n (q-n)/\lambda},$$

where p_n is the probability that exactly n units arrive during a service period.

Deduce the expected waiting time per customer (Weiss, 1979).

10.19 Consider a k out of n-system, each of whose components has an exponential life time distribution with mean $(1/a)$. Show that the variance of the life time of the system is

$$(1/a^2) \sum_{r=k}^{n} (1/r^2).$$

10.20 The lifetime of the ith component of an n-component system has an uniform distribution in $(0, t)$. Find the reliability of the system when the components are connected (i) in series, and (ii) in parallel.

10.21 Consider an equipment consisting of two components connected in parallel. Each component has a failure rate a when both the components are operative and has a higher failure rate ca $(c > 1)$ when either of the components has failed. Find the reliability of the system, and also its mean lifetime.

10.22 Two-state process (see also Example 5 (b), Ch. 4). Consider a system consisting of a single component that can be in two states, functioning and non functioning. Suppose that the failure and the repair processes are both Poissonian with rates a and b respectively. Find the probability that the system is in an acceptable state at the instant t given that the system was fully operative at $t = 0$.

This probability indicates the *availability* of the system, and is denoted by $A(t)$.

10.23 **System reliability in a random environment**

Consider a parallel system of n components, situated in a random environment such that

(i) shocks occur according to a Poisson process with parameter λ,

(ii) *when* a shock occurs, each component fails, independently of the others, with probability $p(= 1 - q)$.

If the interval Z upto the first failure of the entire system has d.f. $H(t) = \Pr\{Z \leqslant t\}$, then show that

$$\overline{H}(t) = 1 - H(t) = \sum_{r=1}^{n} (-1)^{r-1} \binom{n}{r} \exp\{-\lambda(1 - q^r)\,t\}.$$

Suppose that the condition (i) is replaced by (ia) :

(ia) shocks occur according to a renewal process having interarrival distribution with d.f. $F(x)$, and with mean $1/\lambda$.

If $\overline{H}^*(s) = \int_0^\infty e^{-st}\{1 - H(t)\}\,dt$ and $f^*(s) = \int_0^\infty e^{-st}\,dF(t)$, then

show that

$$\overline{H}^*(s) = \sum_{r=1}^{n} (-1)^{r-1} \binom{n}{r} \left\{\frac{1 - f^*(s)}{s[1 - q^r f^*(s)]}\right\}.$$

Show that the mean time μ_n until failure of the system is

$$\mu_n = (1/\lambda) \sum_{r=1}^{n} (-1)^{r-1} \binom{n}{r} \Big/ (1 - q^r).$$

(Rade, 1976).

10.24 Consider the same situation as described in Exercise 23 but subject to the following replacement rule: The system is exchanged before failure (as a preventive measure before failure) if the total number of failed components exceeds k and is replaced upon failure iff all the components fail and otherwise it is left alone. A cost c_1 is incurred when the failed system is replaced and a cost $c_2 (< c_1)$ is incurred when the system is exchanged before failure.

Show that the meantime that the number of failed components exceeds k for the first time is

$$\frac{1}{\lambda} \sum_{r=0}^{k} (-1)^r \binom{n}{r} \sum_{i=0}^{k-r} \left\{ \binom{n-r}{i} \Big/ (1 - q^{n-i}) \right\}.$$

Verify that this is equal to the meantime to failure of a $(k+1)$-out-of-n system.

Show that the expected cost per unit of time for an infinite time span in the steady-state is

$$C(k) = \frac{c_2 + (c_1 - c_2) \sum_{r=0}^{k} (-1)^r \binom{n}{r} p^{n-r} \left[\sum_{i=0}^{r} \left\{ \binom{r}{i} (-1)^i / (1 - q^{n-i}) \right\} \right]}{(1/\lambda) \sum_{r=0}^{k} (-1)^r \binom{n}{r} \left[\sum_{i=0}^{k-r} \left\{ \binom{n-r}{i} \Big/ (1 - q^{n-i}) \right\} \right]},$$

$k = 0, 1, \ldots, n-1.$

(Nakagawa, 1979a).

10.25 Consider a parallel system of $n (\geqslant 2)$ components in a random environment which generates shocks at a mean interval of $(1/\lambda)$. Suppose that the damage caused to a component is cumulative and the probability that a component fails by ith shock is $p(i), i = 1, 2, \ldots$. If $P(j) = \sum_{i=1}^{j} p(i)$ and $\bar{P}(j) = 1 - P(j)$, then show that the mean time until failure of the system is

$$\mu_n = \left(\frac{1}{\lambda}\right) \sum_{r=1}^{n} (-1)^{r-1} \binom{n}{r} \left\{ \sum_{j=0}^{\infty} [\bar{P}(j)]^r \right\}.$$

Obtain μ_n (of Exercise 23) as a particular case from the above. [Hint: put $p(i) = q^{i-1} p, i = 1, 2, \ldots$] (Nakagawa, 1979b).

REFERENCES

K.L. Arora, "Two server bulk-service Queueing Processes," *Opns. Res.*, **12**, 286–294 (1964).

Abdel-Hameed, and F. Proschan, "Nonstationary Shock Models," *Stoch. Proc. Appl.*, **1**, 383–404 (1973).

———, "Shock Models with underlying Birth Process," *J. Appl. Prob.*, **12**, 18–28 (1975).

N.T.J. Bailey, "On queueing processes with bulk service," *J.R.S.S.B.*, **16**, (1954) 80–87.

———, "A continuous time treatment of a simple Queue using generating functions," *J.R.S.S.B.*, **16**, 288–291 (1954).

A.A. Balkema and L. De Haan, "Residual life time at great age," *Ann. Prob.*, **2**, 792–804 (1974).

D.Y. Barrer, "Queueing with impatient customers and indifferent clerks," *Opns. Res.*, **5**, 650–656 (1957).

R.E. Barlow and F. Proschan, *Mathematical Theory of Reliability*, Wiley, New York (1965).

———, *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart, New York (1975).

———, "Some current academic research in system Reliability theory," *IEEE Trans. Reliability*, **R-25**, 3, 198–202 (1976).

B.V. Gnedenko, Yu. K Belyayev and A.D. Solovyev, *Mathematical Methods of Reliability Theory*, Academic, New York, (1969).

E. Borél, "Sur l'emploi du théorème de Bernoulli pour faciliter le calcul d'une infinite de coefficients. Application au problème de l'attente à un guichet," *Comptes Rendus*, *Acad. Sci.*, Paris, **214**, 452–456 (1942).

U.N. Bhat, "Imbedded Markov chain analysis of bulk queues," *J. Aust. Math. Soc.*, **4**, 244–263 (1963).

———, *A Study of the Queueing Systems M/G/1 and GI/M/1*, Lecture Notes in Operations Research and Math. Eco., No. 2, Springer-Verlag, (1968).

———, "Sixty years of Queueing Theory," *Management Science*, **15 B**, 280–292 (1969).

Z.W. Birnbaum, J.D. Esary, and A.W. Marshall, "A stochastic characterization of Wear-out for Components and Systems," *Ann. Math. Stat.*, **37**, 816–825 (1966).

A.R. Bloemena, "On a queueing process with a certain type of bulk service," *Bull. Inst. Statist.*, **37**, 219–227 (1960).

A. Borthakur, "On Busy period of a Bulk Queueing System with a general Bulk service rule," *Opsearch*, **12**, 40–46 (1975).

———, "A Study of some Queueing Models," Ph. D. thesis, Gauhati University, India (1975).

A. Borthakur and J. Medhi, "A queueing system with arrivals and service in batches of variable size," *Cah. du Centre d'etudes de Rech.*

Oper., **16**, 117–126 (1974).

E. Brockmeyer, H.L. Halstorm, and A. Jenson, *The Life and Works of A.K. Erlang*, (Transl. of the Danish Academy of Sciences) No. **2** (1948).

S.L. Brumelle, "A generalisation of $L = \lambda W$ to moments of queue length and waiting times," *Opns. Res.*, **20**, 1127–1136 (1972).

———, "A generalization of Erlang's loss system to state dependent arrival and service rates," *Math. of Opns. Res.*, **3**, 10–16 (1978).

P.J. Burke, "The output of a Queueing system," *Opns. Res.*, **4**, 699–704 (1956).

———, "The output process of a stationary $M/M/s$ Queueing system," *Ann. Math. Stat.*, **39**, 1141–1152 (1968).

D.C. Champernowne, "An elementary method of solution of the Queueing problem with a single server and a constant parameter," *J.R.S.S.B.*, **18**, 125–128 (1956).

M.L. Chaudhry and J.G.C. Templeton, "The queueing system $M/G^B/1$ and its ramifications," *Euro. J. Opns. Res.*, **6**, 56-60 (1981).

A.B. Clarke, "Maximum likelihood estimates in a simple Queue," *Ann. Math. Stat.*, **28**, 1036–1040 (1957).

———, Ed., *Mathematical Methods in Queueing Theory*, Lecture Notes in Eco. and Math. Syst., No. **98**, Springer-Verlag, New York, (1974).

J.W. Cohen, *The Single Server Queue*, North Holland, Amsterdam, (1969).

B.W. Conolly, "A difference equation technique applied to the simple Queue with arbitrary arrival interval distribution," *J.R.S.S.B.*, **21**, 168–175 (1958).

———, *Lecture Notes on Queueing Systems*, E. Horwood, Chichester, (1975).

B.W. Conolly and J. Chan, "Generalised birth and death queueing processes; recent results," *Adv. in Appl. Prob.*, **9**, 125–140 (1977).

D.R. Cox, "Some problems of statistical analysis connected with congestion," in *Proc. Symp on Queue Th.*, University of N. Carolina, Chapel Hill, (1965).

T.B. Crabill, D. Gross, and M.G. Magazine, "A classified bibliography of research on optimal design and control of queues," *Opns. Res.*; **25**, 219–232 (1977).

D.J. Daley, "Queueing output processes," *Adv. Appl. Prob.*, **8**, 395–415 (1976).

F. Downton, "Waiting time in bulk service Queues," *J.R.S.S.B.*, **17**, 256–261 (1955).

———, "On limiting distributions arising in bulk service," *J.R.S.S.B.*, **18**, 265–274 (1956).

E.G. Enns, "The trivariate distribution of the maximum queue length, the number of customers served and the duration of the busy period for the $M/G/1$ Queueing system," *J. Appl. Prob.*, **6**, 154–161 (1969).

J.D. Esary, A.W. Marshall, and F. Proschan, "Shock models and wear Processes," *Ann. Prob.*, **1**, 627–649 (1973).

A.T. Fabens, "The solution of Queueing and Inventory models by semi-Markov processes," *J.R.S.S.B.*, **23**, 113–127 (1961).

R. Fortet, "Random distributions with an application to Telephone Engineering," *Proc. Berk. Symp. Math. Stat. and Prob.*, Los Angeles, **2**, 81–88 (1956).

P.M. Ghare, "Multichannel queueing system with bulk service," *Opns. Res.*, **16**, 189–192 (1968).

B.V. Gnedenko, and I.N. Kovalenko, *Introduction to Queueing Theory*, Israel Program for Scient. Tran. Ltd., Jerusalem, (1968).

D. Gross, and C.M. Harris, *Fundamentals of Queueing Theory*, Wiley, New York, (1974).

R.R.P. Jackson and P. Aspden, "A transient solution to the multistage Poisson queueing system with infinite servers," *Opns. Res.*, **28**, 618–622 (1980).

N.K. Jaiswal, "The Time-dependent solution of the bulk-service queueing problem," *Opns. Res.*, **8**, 773–781 (1960).

———, *Priority Queues*, Academic, New York (1968).

S. Karlin, and J.L. McGregor, "The Classification of Birth and Death Processes," *Trans. Ann. Math. Soc.*, **86**, 366–400 (1957).

J.S. Kaufman, "The busy probability in *M/G/N-N* loss systems," *Opns. Res.*, **27**, 204–206 (1979).

J. Keilson, and A. Kooharian, "On Time-dependent Queueing Processess," *Ann. Math. Stat.*, **31**, 104–112 (1960).

D.G. Kendall, "Some Problems in the Theory of Queues," *J.R.S.S.B.*, **13**, 151-185 (1951).

———, "Stochastic Processes occurring in the Theory of Queues and their Analysis by the method of Markov chains," *Ann. Math. Stat.*, **24**, 338–354 (1953).

J.F.C. Kingman, "The single server queue in heavy traffic," *Proc. Camb. Phil. Soc.*, **57**, 902-904 (1961).

———, "Poisson counts for random sequence of events," *Ann. Math. Stat.*, **34**, 1217–1232 (1963).

L. Kleinrock, "A conservation law for a wide class of Queueing disciplines," *Nav. Res. Log. Quart.* **12**, 181-192 (1965).

———*Queueing Systems*, Vol. I, Wiley, New York (1975).

L. Kosten, *Stochastic Theory of Service Systems*, Pergamon, New York, (1973).

M. Krakowski, "Conservation methods in Queueing Theory," *Rev. Fran. d'auto. Infor. Rech. Oper.* (R.A.I.R.O.) 7e année, V-I, 3-20 (1973).

A. Lee, *Applied Queueing Theory*, Macmillan, Toronto, (1966).

A.J. Lemoine, "On random walks and stable GI/G/I queues," *Math. Opns. Res.*, **1**, 159-164 (1976).

D. Levhari, and E. Shesinski, "A micro-economic production function." *Econometrika*, **38**, 559-573 (1970).

C.H. Lie, C.L. Hwang, and F.A. Tillman, "Availability of Maintained systems: A state of the art Survey," *A. I. I. E. Trans.*, **9**, 247-259 (1977).

J.D.C. Little, "A proof for the Queueing Formula $L = \lambda W$," *Opns. Res.* **9**, 383-387 (1961).

A.W. Marshall and M. Shaked, " Multivariate shock models for distributions with increasing hazard rate average," *Ann. Prob.*, **7**, 343-358 (1979).

J. Medhi, "Waiting time distribution in a Poisson queue with a general bulk service rule," *Management Science*, **21**, 777-782 (1975).

———"Further results on the waiting time distribution of a Poisson queue with a general bulk service rule," *Cah. du Centre d'etudes de Rech. Oper.*, **21**, 183-191 (1979).

———"Applied Probability and Stochastic Processes: Some areas of application," Part III of the Proceedings of the 66th session, Indian Science Congress Association, 17-32 (1979).

J. Medhi, and A. Borthakur, "On a two server Markovian queue with a general bulk services rule," *Cah. du Centre d'etudes de Rech. Oper.*, **14**, 151-158 (1972).

E.C. Molina, *Poisson's Exponential Binomial Limit*, Van Nostrand, New York, (1942).

P.M. Morse, *Queues, Inventories and Maintenance*, Wiley, New York, (1958).

T. Nakagawa, "Replacement problem of a parallel system in random environment," *J. Appl. Prob.*, **16**, 203-205 (1979a).

———"Further results of replacement problem of a parallel system in random environment," *J. Appl. Prob.*, **16**, 923-926 (1979b).

P. Naor, "On machine interference," *J.R.S.S.B*, **18**, 280-287 (1956).

B. Natvig, "On the input and output processes for a general birth and death queueing model," *Adv. Appl. Prob.* **7**, 576-592 (1975).

M.F. Neuts, "The single server Queue with Poisson input and semi-Markov service times," *J. Appl, Prob.*, **3**, 202-230 (1966).

———, "A general class of bulk queues with Poisson input," *Ann. Math. Stat.*, **38**, 759-770 (1967).

———"Markov renewal branching process," in A. B., Clarke, *Op. cit.* (1974).

———"Algorithmic methods in Probability theory," *TIMS Studies in Man. Sc.*, **7**, North-Holland (1977).

———,"Queues solvable without Rouché's theorem," *Opns. Res.*, **27**, 767-781 (1979).

N.U. Prabhu, "Some results for the Queue with Poisson arrivals," *J. R. S.S.B.*, **22**, 104-107 (1960).

———, *Queues and Inventories*, Wiley, York, (1965).

———, "Ladder variables for continuous time stochastic processes," *Z. Wahr. verw. Gleb* , **16**, 157-164 (1970).

———, "Limit theorems for the single server queue with traffic intensity one," *J. Appl. Prob.*, 7, 227-233 (1970).

———"Wiener-Hopf techniques in Queueing theory," in A.B. Clarke, *Op. cit.* (1974).

N.U. Prabhu and U.N. Bhat, "Further results for the Queue with Poisson arrivals," *Opns. Res.*, **11**, 380-386 (1963).

N.U. Prabhu and M. Rubinovitch, "On a continuous time extension of Feller's lemma," *Z. Wahr. verb. Gieb.*, **17**, 220-226 (1971).

L. Rade, "Reliability systems in random environment," *J. Appl. Prob.*, **13**, 407-410 (1976).

E. Reich, "Waiting times when Queues are in tandem," *Ann. Math. Stat.*, **28**, 768-773 (1957).

J. T. Runnenberg, "Probabilistic interpretation of some formula in Queueing theory," *Bull d'Inst. Inter de Stat.*, **37**, 405-414 (1960).

M. Scott and M.B. Ulmer Jr., "Some results for a simple queue with limited waiting room," *Zeit. f. Opns. Res.*, **16**, 199-204 (1972).

W. Shonick, "A stochastic model for occupancy-related random variables in general acute hospitals," *J. Am. Stat. Ass.*, **65**, 1474-1500 (1970).

G.P. Sphicas, and D.G. Shimshak, "Waiting time variability in some single server queueing systems," *J. Opnl. Res. Soc.*, **29**, 65-70 (1978).

M.J. Sobel, "Optimal operation on Queues," in A.B. Clarke, *Op. cit.* (1974).

S. Stidham, "A last word on $L = \lambda W$," *Opns. Res.*, **22**, 417-421 (1974).

S. Stidham and N.U. Prabhu, "Optimal control in Queueing system," in A.B Clarke, *Op. cit.* (1974).

L. Takacs, *Introduction to the Theory of Queues*, Oxford University Press, Oxford (1962).

———, "On Erlang's Formula," *Ann. Math. Stat.*, **40**, 71-78 (1969).

J. Teghem, J. Loris-Teghem and J.P. Lambotte, *Modeles d'attente M/G/1 et GI/M/1 à Arrivèes et Services en Groupes*, Lecture Notes on O.R., **8**, Springer-Verlag, Berlin, (1969).

F.A. Tillman, C.L. Wang, and W. Kuo, "Optimisation techniques for system reliability with redundancy: a review," *IEEE Trans. on Rel.*, **R-26**, 148-155 (1977).

H.J. Weiss, "The computation of optimal control limits for a queue with batch service," *Management Science*, **25**, 320-328 (1979).

W. Whitt, 'Heavy traffic limit theorems for queues: A survey," in A.B. Clarke, *Op. cit.* (1974).

R.W. Wolff, "Problems of statistical inference for birth and death queueing models," *Opns. Res.*, 13, 343-357 (1965).

ANSWERS TO EXERCISES

Chapter 1

1.1 (a) $s(1-s-s^2)^{-1}$.

1.4 rq/p, rq/p^2; $\{p/(1-qs)\}^{mr}$, mrq/p, mrq/p^2.

1.5 (b) (i) $(e^a-1)^{-2}(e^{as}-1)^2$,

(ii) $(e^{a_1}-1)^{-1}(e^{a_2}-1)^{-1}(e^{a_1 s}-1)(e^{a_2 s}-1)$

(c) $(e^a-1)^{-n}(e^{as}-1)^n$;

$$\Pr(S_n=m)=[(e^a-1)^{-n}]\sum_{r=0}^{n-1}(-1)^r\binom{n}{r}\frac{a^m(n-r)^m}{m!}$$

$$=\frac{n!\,a^m S(m,n)}{m!\,(e^a-1)^n}\quad \text{for } m=n,\, n+1,\, \ldots,$$

where $S(m,n)=\sum_{r=1}^{n}(-1)^{n-r}\binom{n}{r}r^m/n!$ is the Stirling number of the second kind with arguments m and n.

$$E(S_n)=na\,e^a(e^a-1)^{-1}.$$

1.7 (a) (i) $\exp\{(\lambda p/(1-qs)-1\}$.

(ii) $E(S_N)=aq/p$; $\operatorname{var}(S_N)=(aq/p^2)+(aq^2/p^2)$.

1.9 $P(s_1,s_2)=\dfrac{p_1p_2}{(1-q_1s_1)(1-q_2s_2)}$.

1.10 (i) $\dfrac{a}{s^2+a^2}$, (ii) $\dfrac{a}{s+2}\,(s>-2)$

(iii) $\dfrac{s^4+4s^2+24}{s^3}(s>0)$, (iv) $\dfrac{6}{(s+3)^4}$

(v) $\dfrac{s+1}{s^2+2s+5}$, (vi) $\dfrac{\Gamma(n+1)}{(s-a)^{n+1}}$

(vii) $\dfrac{a^n}{(s+a)^{n+1}}$, (viii) $e^{-a\sqrt{s}}$

(ix) $\exp(1/s)/s^{r+1}$, (x) $\exp\left(\dfrac{-s}{s+1}\right)/(s+1)^{r+1}$

(xi) $\dfrac{\exp\{a\lambda/(\lambda+s)\}-1}{\exp(a)-1}$.

1.12 (i) $\lambda k e^{-\lambda x}$, (ii) $E(X)=k/\lambda$, (iii) $\lambda k/(s+\lambda)$.

1.14 $(1-e^{-s})^2/s^2$.

1.15 (i) $C_1 + C_2(2)^n$; (ii) $C_1 + C_2(-1)^n$;
(iii) $C_1 + C_2 n + C_3 2^n$.

1.16 $\dfrac{(q/p)^a - (q/p)^n}{(q/p)^a - 1}$; $1 - n/a$.

1.21 Three eigenvalues are 1, $(-1 \pm \sqrt{5})/4$; the corresponding left eigenvectors are (1, 1, 1); $((-1-\sqrt{5})/4, (-1+\sqrt{5})/4, 1)$; $((-1+\sqrt{5})/4, (-1-\sqrt{5})/4, 1)$.

1.22 (i) Imprimitive, (ii) Imprimitive, (iii) Primitive, (iv) Imprimitive, (v) Imprimitive.

Chapter 3

3.1 $p_{00} = p_{11} = p,\ p_{01} = p_{10} = q;$

$$p_{00}^{(2)} = p_{11}^{(2)} = p^2 + q^2,\ p_{01}^{(2)} = p_{10}^{(2)} = 2pq.$$

$$p_{00}^{(3)} = p_{11}^{(3)} = p(p^2 + 3q^2),\ p_{01}^{(3)} = p_{10}^{(3)} = q(3p^2 + q^2).$$

$$\Pr\{X_2 = 0 \mid X_0 = 1\} = p_{10}^{(2)},\ \Pr\{X_3 = 1 \mid X_0 = 0\} = p_{01}^{(3)}.$$

3.2 $p_{jj} = 0,\ p_{jk} = \frac{1}{2},\ k \neq j$

$$\Pr\{X_2 = 1 \mid X_0 = 1\} = p_{11}^{(2)} = \tfrac{1}{4},\ \Pr\{X_2 = 2 \mid X_0 = 3\} = \tfrac{3}{8}$$

$$\Pr\{X_2 = 3 \mid X_0 = 2\} = \tfrac{3}{8}.$$

When the number of children is $m\ (\geqslant 3)$, P will be an $m \times m$ matrix such that $p_{jj} = 0,\ p_{jk} = 1/(m-1),\ j \neq k$.

3.3
$$P = \begin{pmatrix} p & q & 0 & 0 \\ 0 & 0 & p & q \\ p & q & 0 & 0 \\ 0 & 0 & p & q \end{pmatrix};\ P^m = (p^2, pq, pq, q^2),\ m = 2, 3, \ldots.$$

3.4 P is a (6×6) triangular matrix (p_{jk}) such that

$$p_{jj} = j/6,\ p_{jk} = 0,\ k < j,\ p_{jk} = 1/6, \quad k > j.$$

For P^2, $p_{jj}^{(2)} = (j/6)^2,\ p_{jk}^{(2)} = \sum_r p_{jr}\, p_{rk} = 0, \qquad k < j$

$$p_{jk}^{(2)} = \sum_r p_{jr}\, p_{rk} = (2k-1)/36$$

$$\Pr\{X_3 = 6\} = 91/216.$$

3.5 $P = (p_{jk})$ is a (3×3) matrix such that $p_{jj} = j/3$,
$p_{j,j+1} = (3-j)/3,\ p_{j,k} = 0,\ k \neq j, j+1.$

The eigenvalues of P are $t_j = j/3$, $j = 1, 2, 3$. Left eigenvectors corresponding to these eigenvalues are $(1, 0, 0)$, $(1, \frac{1}{2}, 0)$ and $(1, 1, 1)$ respectively; the right eigenvectors are $(1, -2, 1)'$, $(0, 1, -1)'$ and $(0, 0, 1)'$ respectively, so that $c_1 = 1$, $c_2 = 2$, $c_3 = 1$. Thus

$$P = \sum_{j=1}^{3} c_j t_j B_j, \quad P^n = \sum_{j=1}^{3} c_j t_j^n B_j,$$

where $$B_1 = \begin{pmatrix} 1 & -2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0 & 2 & -2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix},$$

$$B_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\lim P^n \to (0, 0, 1), \quad \text{as} \quad n \to \infty.$$

3.6 For r cells, $P = (p_{jk})$ is a $(r \times r)$ matrix such that

$$p_{jj} = j/r, \; p_{j,j+1} = (r - j)/r, \; p_{jk} = 0, \; k \neq j, j + 1.$$

Elements of P^2 are $(p_{jk}^{(2)})$,

where $$p_{jj}^{(2)} = (j/r)^2, \; p_{j\,j+1}^{(2)} = (r - j)\,(2j + 1)/r^2,$$

$$p_{j\,j+2}^{(2)} = (r - j)\,(r - j - 1)/r^2,$$

$$p_{jk}^{(2)} = 0, \; k \neq j, j + 1, j + 2.$$

The eigenvalues of P are $t_j = j/r$, $j = 1, 2, \ldots, r$.

The right eigenvectors $\mathbf{x}_i$ of P are solutions of the equations

$$(P - t_i\, I)\, \mathbf{x}_i = 0, \quad \mathbf{x}_i = (x_{i1}, \ldots, x_{is}, \ldots, x_{ir})'.$$

The solution of the homogeneous equations (which can be obtained by the method outlined in Jordan,† section 173) is given by

$$x_{is} = \binom{s-1}{i-1} \Big/ \binom{r-1}{i-1}, \quad s \geqslant i$$
$$= 0, \quad s < i$$

Similarly $$y_{is} = (-1)^{i-s} \binom{r-s}{i-s}, \quad s \leqslant i$$
$$= 0, \quad s > i$$

whence $1/c_s = 1 \Big/ \binom{r}{s}$. Thus P and P^n can be written down.

It follows that

$$\lim P^n \to (0, 0, \ldots, 1), \quad \text{as} \quad n \to \infty.$$

The interpretation of the limit is intuitively clear.

---

†*Calculus of Finite Differences*, Chelsea, New York (1950).

3.7 (i)–(iv) Irreducible; (v) Reducible, has two pairs of closed chains.

3.8 $p_{00}^{(2n)} = p\binom{2n}{n} p^n q^n.$

Using Stirling's approximation for $n!$ (n large), we get

$$p_{00}^{(2n)} \simeq (4pq)^n/\sqrt{(n\pi)};$$

when $p = \frac{1}{2}$, $\Sigma p_{00}^{(2n)} = \sum_n \frac{1}{\sqrt{(n\pi)}}$ diverges but $p_{00}^{(2n)} \to 0$.

Hence the state 0 is persistent null if $p = \frac{1}{2}$.

When $p \neq \frac{1}{2}$, $p_{00}^{(2n)} < a^n$, $a = 4pq < 1$

and so $\Sigma p_{00}^{(2n)}$ converges; thus the state 0 is transient. Similarly for the state 1.

3.9 Let $f_{jk}^{(n)}$ be the probability that starting from state j, the state k is reached for the first time at nth step. We have

$$\begin{aligned} f_{00}^{(n)} &= 1 - a_1, & n &= 0 \\ &= a_1 a_2 \ldots a_{n-1}(1 - a_n), & n &= 1, 2, \ldots \end{aligned}$$

$$\begin{aligned} F_{00} = \Sigma f_{00}^{(n)} &= (1 - a_1) + \sum_{n=1}^{\infty} a_1 \ldots a_{n-1}(1 - a_n) \\ &= 1 - \lim (a_1 \ldots a_n). \end{aligned}$$

Thus the state 0 is recurrent iff $\lim (a_1 \ldots a_n) = 0$.
We have

$$\mu_{00} = \Sigma (n + 1) f_{00}^{(n)} = 1 + \sum_{n=1}^{\infty} a_1 \ldots a_n.$$

Hence the state 0 is positive iff $\sum_{n=1} a_1 \ldots a_n$ is finite.

The other states 1, 2,... will be recurrent (or persistent) and non-null (or null) according as the state 0 is.

3.10 To proceed as in 3.9.

Chapter 4

4.3 $\exp\{\lambda t(s - 1)/(1 - qs)\}$; $\lambda t/p$, $\lambda t(1 + q)/p^2$.

4.4 28; 42

4.7 $p_n'(t) = -n\mu p_n(t) + (n + 1)\mu p_{n+1}(t)$

$E\{N(t)\} = i \exp(-\lambda t)$

$\operatorname{var}\{N(t)\} = i \exp(-\lambda t)\{1 - \exp(-\lambda t)\}$.

4.9 1, when $\mu > \lambda$ and $\left(\frac{\mu}{\lambda}\right)^i$, when $\lambda > \mu$.

4.10 $p_n'(t) = -\{(\lambda + \mu)n + a\} p_n(t) + \{\lambda(n-1) + a\} p_{n-1}(t) + (n + 1)\mu p_{n+1}(t),$
$n \geqslant 1$

$p_0'(t) = -ap_0(t) + \mu p_1(t)$

As t tends to ∞, $M(t)$ tends to $a/(\mu - \lambda)$ for $\mu > \lambda$.

4.12 $E(T_n) = \left(\frac{1}{\lambda}\right) \sum_{k=1}^{n} (-1)^{k+1} \binom{n-1}{k} \frac{1}{k} = \frac{1}{\lambda} \sum_{k=1}^{n-1} \frac{1}{k}$.

4.13 $E(Z_k) = kt/(n+1)$.

4.14 (i) 0.458, (ii) 0.956.

4.16 $\Pr\{N = n\} = \int_0^\infty e^{-\lambda t}(1 - e^{-\lambda t})^n \mu e^{-\mu t}\, dt = \frac{\mu}{\lambda}\, \beta\left(n+1, \frac{\mu}{\lambda} + 1\right)$

(Note that N does not include the parent.)

4.18 Z is negative exponential with mean $1/\Sigma \lambda_i$, for $i = 1, 2, \ldots n$.

4.19 When $\Sigma \alpha_j < \infty$, $\{v_j\}$ defines a probability distribution.

When $\Sigma \alpha_j = \infty$, $v_j = 0$ for all j.

4.22

$$A = \begin{bmatrix} -2a & a & a & 0 \\ b_2 & -(a+b_2) & 0 & a \\ b_1 & 0 & -(a+b_1) & a \\ 0 & b_1 & b_2 & -(b_1+b_2) \end{bmatrix}$$

$$V = (kb_1b_2,\ kab_1,,\ kab_2\ ka^2),$$

where $k^{-1} = (a+b_1)(a+b_2)$.

4.23 $v_0 = k$, $v_1 = (a/b)\,k$, $v_2 = (a/b)^2 k$, where

$k = [1 + (a/b) + (a/b)^2]^{-1}$; reqd. prob. $= v_0 + v_1$.

Chapter 6

6.4
$$L^*(s) = \{1/(s+B)^2\}\,[A + \{ab + A(2A+B)\}/s + \{ab(4A+B)\}/s^2 + 2(ab)^2/s^3],$$

$A = pa + (1-p)b$, $B = (1-p)\,a + pb$.

$$L(t) = \sum_{i=1}^{4} K_i a_i(t),$$

where

$K_1 = A$, $K_2 = ab + A(2A+B)$, $K_3 = ab\,(4A+B)$, $K_4 = 2(ab)^2$

$a_1 = te^{-Bt}$, $a_2 = (1 - e^{-Bt})/B^2 - te^{-Bt}/B$,

$a_3 = t(1 + e^{-Bt})/B^2 - 2(1 - e^{-Bt})/B^3$,

$a_4 = (t^2/2B^2) - t(2 + e^{-Bt})/B^3 + 3(1 - e^{-Bt})/B^4$.

6.13 $M_1(t) = abt/(a+b) + \{a^2/(a+b)^2\}\{1 - \exp(-(a+b)\,t)\}$

$M_0(t) = abt/(a+b) + \{ab/(a+b)^2\}\{\exp(-(a+b)\,t) - 1\}$.

6.15 $\dfrac{AM(T)+B}{T} \to \dfrac{A}{\mu} + \dfrac{1}{T}\left\{B + A\,\dfrac{\sigma^2 - \mu^2}{2\mu^2}\right\} + o(T^{-1})$

(μ, σ^2 are the mean and variance of the lifetime distribution),

Chapter 7

7.1 (i) $f_{jk}^{*}(s)$ can be found from

$$F^{*}(s) = [I - C^{*}(s)]^{-1} A^{*}(s);$$

$$f_{11}^{*}(s) = \frac{(3s+7)(s+2)^{2}}{s\Delta_1}$$

$$f_{12}^{*}(s) = \frac{3(s+1)(s+3)}{s\Delta_1}$$

$$f_{21}^{*}(s) = \frac{(3s+7)(s+4)}{s\Delta_1}$$

$$f_{22}^{*}(s) = \frac{(3s+1)(s+3)^{2}}{s\Delta_1}$$

where

$$\Delta_1 = 3s^3 + 22s^2 + 52s + 37.$$

On factorising Δ_1, then splitting into partial fractions, and finally inverting the transforms, expressions for $f_{ij}(t)$ can be obtained.

(ii) $f_1 = 28/37$, $f_2 = 9/37$.

7.2 (i) $g_{11}(t) = (2/3)e^{-t} + e^{-3t} + 2te^{-2t} - e^{-2t}$

$g_{12}(t) = e^{-2t} + 2te^{-2t}$

$g_{21}(t) = (3/4)e^{-3t} + (1/4)e^{-t/3}$

$$g_{22}(t) = 2e^{-4t} + \left\{\frac{3e^{-2t}}{5} + \frac{3e^{-t/3}}{20} - \frac{3e^{-3t}}{4}\right\}.$$

Using (5.6), expressions for $g_{ij}^{*}(s)$ can be found.

Note that $f_{ij}^{*}(s)$ can be found also by using (5.4) and (5.4a).

(ii) $\mu_{11} = 37/36$, $\mu_{22} = 37/24$

(iii) $\mu_{12} = 7/3$, $\mu_{21} = 3/4$.

7.3 (ii) $\mu_{01} = 7/6$, $\mu_{10} = 23/7$

(iii) $f_0 = 49/187$, $f_1 = 138/187$.

7.4 The relation can be found by simple probabilistic argument. Taking L.T. of the relation, and multiplying both sides by s,

$$\{sf_{ij}^{*}(s)\} = \{se_{ij}^{*}(s)\}\{h_j^{*}(s)\},$$

and then taking limits, as $s \to 0$, one gets the relation $e_j = f_j/M_j$.

7.6 $\operatorname{var}(V_{12}) = 17/9$, $\operatorname{var}(V_{21}) = 13/16$.

7.10 $Q_{i,j}(t) = 0, \qquad i \geqslant 1, j < i-1$

$$= \int_0^t \frac{e^{-ay}(ay)^{j-i+1}}{(j-i+1)!}\, dB(y), \qquad i \geqslant 1, j \geqslant i-1$$

$$Q_{0,j}(t) = \int_0^t \{1 - e^{-a(t-y)}\} \frac{e^{-ay}(ay)^j}{j!}\, dB(y), \quad j \geqslant 0$$

7.11
$$Q_{0,j}(t) = \int_0^t \{1 - e^{-a(t-y)}\} \frac{e^{-ay}(ay)^j}{j!}\, dB(y), \quad j \geqslant 0$$

$$\begin{aligned} Q_{i,j}(t) &= \int_0^t e^{-ay} \frac{(ay)^j}{j!}\, dB(y), && 1 \leqslant i \leqslant c,\ j \geqslant 0 \\ &= 0, && i \geqslant c,\ j < i - c \\ &= \int_0^t \frac{e^{-ay}(ay)^{j-i+c}}{(j-i+c)!}\, dB(y), && i \geqslant c,\ j \geqslant i - c. \end{aligned}$$

7.12 Equality holds in the former case and does not hold in case of the latter.

Chapter 9

9.2 $(\sqrt{5} - 1)/2$

9.6 $n^{r-1}/(n+1)^r,\ r = 1, 2, \ldots$

9.9
$$H(s) = \frac{1 - \sqrt{(1 - 4pq\,s)}}{2q}$$

9.10 $b_k = (1 - q/p)(q/p)^{k-1},\ k = 1, 2, \ldots$

$$B(s) = \sum_{k=1}^{\infty} b_k s^k = \frac{(p-q)\,s}{p - qs}$$

9.12 (i)
$$\begin{aligned} r_n &= \frac{p^n\, q^{n-1}(q-p)^2}{(q^{n+1} - p^{n+1})(q^n - p^n)},\ p \neq q \\ &= 1/n\,(n+1), \qquad p = q, \qquad n = 1, 2, 3, \ldots. \end{aligned}$$

(ii)
$$\begin{aligned} r_n &= \frac{s_0\,(1 - s_0)\, m^{n-1}\,(m-1)}{(m^{n-1} - s_0)(m^n - s_0)}, \qquad m \neq 1 \\ &= \frac{(1-c)\,c}{\{1 + c\,(n-1)\}\,\{1 + c\,(n-2)\}},\ m = 1,\ n = 1, 2, 3, \ldots. \end{aligned}$$

9.14
$$\begin{aligned} \operatorname{var}(X(t)) &= \frac{u''(1) - u(1)}{u'(t)} \exp\{u'(1)\,t\}\,[\exp\{u'(1)t\} - 1] \\ & \qquad\qquad \text{if } u'(1) \neq 0 \\ &= u''(1)t \qquad\qquad \text{if } u'(1) = 0. \end{aligned}$$

APPENDIX 1

ABBREVIATIONS

$\mathrm{Pr}(A)$; $\mathrm{Pr}\{A\}$	probability of the event A
r.v.	random variable
p.d.f.	probability density function (or density function)
p.g.f.	probability generating function
p.f.	probability mass function
d.f.; c.d.f.	cumulative distribution function
iff	if and only if
i.i.d.	identically and independently distributed
$E(X)$	expectation of the r.v. X
var (X)	variance of the r.v. X
L.T.	Laplace transform
C.K.	Chapman-Kolmogorov
t.p.m.	transition probability matrix
i.t.p.m.	interval transition probability matrix
w.r.t.	with respect to
l.h.s. (r.h.s.)	left (right) hand side
$\sim$	asymptotic to
$X*Y$	convolution of two random variables X and Y
F_1*F_2	convolution of two distributions with d.f. F_1 and F_2
F^{n*}	convolution of n i.i.d. random variables having common d.f. F
▲	indicates the end of a proof

GLOSSARY OF COMMONLY USED NOTATIONS

Chapter 1

$P(s)$	probability generating function (p.g.f.) of a r.v.
S_N	sum of N i.i.d. random variables X_i, where N is a random variable independent of X_i
$G(s)$	p.g.f. of N
E_k	r.v. having Erlang-k distribution
$f^*(s)$	$= \int_0^\infty e^{-st}\,dF(x)$ (L.T. of a r.v. with d.f. F)
t_i	eigenvalue of a matrix
$p(A)$	spectral radius of the matrix A
$\mathbf{x}\ (\mathbf{x}')$	column (row) vector
$\mathbf{x}$	right eigen vector
$\mathbf{y}'$	left eigen vector
P	stochastic matrix

Chapter 2

$\{X(t)\}$; $\{X_t\}$	stochastic process in continuous time
$\{X_n\}$	stochastic process in discrete time
$C_{s,t}$; $C(s, t)$	covariance function between $X(t)$ and $X(s)$

Chapter 3

p_{jk}	transition probability from state j to state k
$P = (p_{jk})$	t.p.m. with elements p_{jk}
$p_{jk}^{(m)}$	m-step transition probability from state j to state k
$P^{(m)}$	t.p.m. with elements $p_{jk}^{(m)}$
$f_{jk}^{(n)}$	probability that the system reaches the state k for the first time after n transitions, given that it started with state j
F_{jh}	$(= \sum_n f_{jk}^{(n)})$, i.e. probability that starting with state j, the system will ever reach state k
μ_{jk}	expected first passage time from state j to state k
$P_{jj}(s)$, $F_{jj}(s)$	p.g.f. of $\{p_{jj}^{(n)}\}$ and $\{f_{jj}^{(n)}\}$ respectively

t_i, $\mathbf{x}_i$, $\mathbf{y}_i'$	eigenvalue, right and left eigenvectors respectively of P
v_j	$= \lim p_{ij}^{(n)}$ as $n \to \infty$
$V(s)$	p.g.f. of $\{v_j\}$

Chapter 4

$N(t)$	number of occurrences of an event by time t (or upto epoch t)
$p_n(t)$	$\Pr\{N(t) = n \mid N(0) = .\}$
$P(s, t)$	p.g.f. of $\{N(t), t \geqslant 0\}$ or of $\{p_n(t)\}$
$M(t)$	compound Poisson process, i.e. $\sum_{i=1}^{N(t)} X_i$, where X_i, are i.i.d. r.v's and $N(t)$ is a Poisson process
$Q(s, t)$	p.g.f. of a non-homogeneous Poisson process
$\{X(t)\}$	Markov process in continuous time with discrete state space or continuous time Markov chain
$p_{ij}(t)$; $p_{i,j}(t)$	probability of transition from state i to state j by time t
$P(t)$	t.p.m. with elements $p_{ij}(t)$
v_j	$\lim p_{ij}(t)$ as $t \to \infty$
a_{ij}	transition density, i.e. $p'_{ij}(t)\|_{t=0}$
A	rate matrix or transition density matrix whose elements are (a_{ij})

Chapter 5

$\{X(t)\}$	Wiener process
μ_t	mean-value function of $X(t)$
$\sigma^2(t)$	variance function of $X(t)$
$M(T)$	maximum of $X(t)$ in $0 \leqslant t \leqslant T$ $(\max_{0 \leqslant t \leqslant T} X(t))$
T_a	first passage time to a fixed point a of a Wiener process with $X(0) = 0$, $\mu = 0$
$U(t)$	Ornstein-Uhlenbeck process (O-U.P)
$m(t)$	mean-value function of $U(t)$
$\sigma^2(t)$	variance-function of $U(t)$

Chapter 6

X_n	non-negative independent r.v. (interarrival time between $(n-1)$th and nth renewal) having mean μ
F	d.f. of X_n
S_n	nth renewal epoch, i.e. $S_n = \sum_{i=1}^{n} X_i$
$N(t)$	number of renewals in $[0, t]$ $(N(t) = \sup\{n: S_n \leqslant t\})$
$M(t)$	expected number of renewals in $[0, t]$
$M*(s)$	L.T. of $M(t)$

$N_A(t, T)\,(N_B(t, T))$	number of renewals in $[0, t]$ under age (block) replacement policy with replacement interval T
G	d.f. of initial distribution of a modified renewal process
$N_D(t)\,(N_e(t))$	number of renewals in $[0, t]$ for modified (equilibrium) renewal process
$F_e(t)$	equilibrium distribution of F having d.f. $\int_0^\infty \{1 - F(x)\}\,dx/\mu$
$G(t, z)$	p.g.f. of $\{N(t)\}$
$G_D(t, z)\,(G_e(t, z))$	p.g.f. of $N_D(t)\,(N_e(t))$
$G^*(s, z)$	L.T. of $G(t, z)$
$M_i(t)$	renewal function of an alternating (or two-stage) renewal process ($i = 0, 1$)
$Y(t)$	residual or spent lifetime (*forward recurrence time*) at time (epoch) t, i.e. $Y(t) = S_{N(t)+1} - t$
$Z(t)$	age (*backward recurrence time*) at time t, i.e. $Z(t) = t - S_N(t)$
Y_n	reward or cost associated with nth renewal
$\{X_n, Y_n\}$	renewal reward process
$Y(t)$	total (cumulative) reward earned or cost incurred by time t
$C(t)$	expected cumulative reward (or cost) by time t

Chapter 7

$\{X(t)\}$	Markov process
t_n	nth transition epoch
$X_n = X(t_n +)$	state at transition occurring at epoch t_n
τ_n	time interval $(t_{n+1} - t_n)$ between nth and $(n + 1)$st transitions
$\{X_n, t_n\}$	Markov renewal process
Y_t	$(= X_n$ on $t_n \leqslant t < t_{n+1})$, i.e. state of the process at its most recent transition
$Q_{jk}(t);\ Q_{j,k}(t)$	$\Pr\{X_{n+1} = k, \tau_n \leqslant t \mid X_n = j\}$
p_{jk}	$\lim Q_{jk}(t)$ as $t \to \infty$
T_{jk}	(conditional) waiting time in state j given that the next transition is to state k
$W_{jk}(t)$	$(= Q_{jk}(t)/p_{jk})$, i.e. d.f. of T_{jk}
α_{jk}	expectation of T_{jk}
$\alpha_{jk}^{(r)}$	rth moment of T_{jk}

T_j	unconditional waiting time in state j
h_j	$\Pr\{T_j > t\}$
M_j	expectation of T_j
$M_j^{(r)}$	rth moment of T_j
S_n^j	successive epochs t_n for which $X_n = j$
$N_j(t)$	number of transitions to state j (renewals of state j) in $[0, t]$
$\mathbf{N}(t)$	$= (N_1(t), N_2(t), \ldots)$
V_{ij}	first passage time into state j from state i
G_{ij}	d.f. of V_{ij}
μ_{ij}	mean first passage time into state j from state i
$f_{ij}(t)$	conditional probability that state at epoch t is j given that initial state was i
$Q_{ij}^{(n)}(t)$	$= \Pr\{X_n = j, t_n \leqslant t \mid X_0 = i\}$
$M(i, j, t)$	Markov renewal function ($= E\{N_j(t) \mid X_0 = i\}$)
F	interval transition probability matrix (i.t.p.m) having elements $f_{ij}(t)$
$f_{ij};\ f_j$	limit $f_{ij}(t)$ as $t \to \infty$
$f_{ij}{*}(s)$	L.T. of $f_{ij}(t)$
$\{v_j\}$	stationary distribution of embedded Markov chain $\{X_n\}$
$U(t)$	($= t - t_n$) time since last transition (backward recurrence time)
$V(t)$	($= t_{n+1} - t_n$) time to next transition (forward recurrence time)

Chapter 8

X_t	wide-sense or covariance stationary process in discrete time
C_k	covariance function of X_t
ρ_k	correlation function of X_t
e_t	purely random process
B	backward shift operator (i.e. $BX_t = X_{t-1}$)
$f(.)$	spectral density function of X_t
$C(z)$	autocovariance generating function of X_t
$S(z)$	autocorrelation generating function of X_t
$X(t)$	stationary process in continuous time
$R(v)\,(\rho(v))$	covariance (correlation) function of $X(t)$

Chapter 9

$\{p_k\}$	offspring distribution
ζ_r	i.i.d. random variable with distribution $\{p_k\}$
$P(s)$	p.g.f. of offspring distribution $\{p_k\}$ and also of r.v. ζ_r, $r = 1, 2, \ldots$
m, σ^2	mean and variance respectively of $\{p_k\}$
X_n	size of nth generation, $n = 1, 2, \ldots$
X_0	size of 0-th generation or ancestor
$\{X_n, n \geqslant 0\}$	Galton-Watson (G.W.) process
$P_n(s)$	p.g.f. of $\{X_n\}$
q	probability of extinction
Y_n	total number of progeny upto nth generation including the ancestor, i.e. $Y_n = 1 + Z_n$
$G_n(s)$	p.g.f. of $\{Z_n\}$, $Z_n = X_1 + \ldots + X_n$
$R_n(s)$	p.g.f. of $\{Y_n\}$
$H(s)$	$\lim R_n(s)$, as $n \to \infty$
b_{kn}	$\Pr\{X_n = k \mid X_n > 0\}$ for a subscritical G.W. process
b_k	$\lim b_{kn}$, as $n \to \infty$
$B_n(s)$	p.g.f. of $\{b_{kn}\}$
$B(s)$	p.g.f. of $\{b_k\}$
$\varphi(0)$	$\lim \Pr\{X_n > 0\}/m^n$, as $n \to \infty$
G.W.I. Process	Galton-Watson process with immigration
$h(s)$	p.g.f. of immigrant distribution
a	mean of immigrant distribution
$X_{(n)}$	size of nth generation of a G.W.I. process
$P_n(s)$	p.g.f. of $\{X_{(n)}\}$
$F(s)$	$\lim P_n(s)$, as $n \to \infty$
$p^{(i)}(r_1, r_2, \ldots r_k)$	probability that an individual of type i produces r_j offspring of type j, $j = 1, 2, \ldots k$ (in case of a multitype G.W. process)
$\mathbf{X}_n = (X_n^{(1)}, \ldots X_n^{(k)})$	population size of k types in the nth generation of a multitype G.W. process
m_{ij}	expected number of offspring of type j produced by an object of type i
$X(t)$	continuous time branching process
$p_{ij}(t)$	$\Pr\{X(t) = j \mid X(0) = i\}$ for a Markov branching process
$F(t, s)$	p.g.f. of $\{p_{ij}(t)\}$
a_j	transition density of $\{X(t),\ t \geqslant 0\}$, $X(0) = 1$

$u(s)$	p.g.f. of $\{a_j\}$
q	$\lim p_{10}(t)$ as $t\to\infty$, i.e. (probability of extinction)
G	d.f. of lifetime of objects in an age-dependent branching process
$F(t, s)$	p.g.f. of an age dependent branching process $\{X(t)\}$ with $X(0)=1$

Chapter 10

λ, μ	mean arrival and service rates respectively
ρ	traffic intensity
$N(t)$	number in the system at epoch t
W_n	waiting time of nth arrival
$W(t)$	virtual waiting time, i.e. duration of time a unit would have to wait were it to arrive at epoch (instant) t
M	exponential (Markovian) distribution
G	arbitrary (general) distribution
$p_n(t)(p_{in}(t))$	probability that the number in the system at t is n given that number at $t=0$ is zero (is i)
p_n	$\lim p_n(t)$ as $t\to\infty$
$L(L_Q)$	expected number in the system (queue)
$W(W_Q)$	expected waiting time in the system (queue)
$w_s(w_q)$	p.d.f. of waiting time in the system (queue)
$f_n(s)$	L.T. of $p_n(t)$
$P(z, t)$	p.g.f. of $\{p_n(t)\}$
T	busy period initiated by a single customer or unit
$N^*(t)$	number in the system at epoch t during a busy period
$M^{(X)}$	arrival streams follow exponential distribution, number arriving at an arrival instant being given by a r.v. X
$M^{a,b}$	bulk service with exponential service time, under a general bulk service rule (i.e., a batch taking a minimum of a units and a maximum of b units)
$p_{1,n}(t)(p_{0,n}(t))$	probability that the number in the queue at time t is n, the service channel being busy (idle) under general bulk service rule
$p_{1,n}\ (p_{0,n})$	limiting value of $p_{1,n}(t)(p_{0,n}(t))$ as $t\to\infty$
M/G/1	
$B(t)\ (b(t))$	d.f. (p.d.f.) of general or arbitrary service time v

$B^*(s)$	$\int_0^\infty e^{-st}\,dB(t)$
t_n	nth departure epoch
$X_n=N(t_n+0)$	number in the system immediately after nth departure
k_r	probability that r units arrive during service time of a unit
p_{ij}	$\Pr\{X_{n+1}=j \mid X_n=i\}$
$\{v_j\}$	stationary distribution of $\{X_n\}$
$V(s)$	p.g.f. of $\{v_j\}$
$K(s)$	p.g.f. of $\{k_r\}$
$W(W_Q(t))$	d.f. of waiting time in the system (queue)
$W^*(s)$	$\int_0^\infty e^{-st}\,dW(t)$

GI/M/1

$A(t)$ $(a(t))$	d.f. (p.d.f.) of general or arbitrary independent interarrival time u
$A^*(s)$	$\int_0^\infty e^{-st}\,dA(t)$
t_n	nth arrival epoch
$Y_n=N(t_n-0)$	number in the system immediately before the arrival of the nth unit
$Z(t)$	$(=Y_n,\ t_n \leqslant t < t_{n+1})$ system size at the most recent arrival
p_{ij}	$=\Pr\{Y_{n+1}=j \mid Y_n=i\}$
g_r	probability that number served in the inter-arrival time between nth and $(n+1)$st units is r
$G(s)$	p.g.f. of $\{g_r\}$
$\{v_j\}$	stationary distribution of $\{Y_n\}$
$r(z)$	$=z-A^*(-\mu z)$
r_0	unique root of $r(z)=0$ lying inside $\lvert z \rvert=1$
$N(t)$	system size at an arbitrary epoch t
$f_{ij}(t)$	$\Pr\{Z(t)=j \mid Z(0)=i\}$
$p_{ij}(t)$	$\Pr\{N(t)=j \mid N(0)=i\}$

Reliability

X_i	Bernoulli r.v. with p.f. $\Pr(X_i=1)=p_i$, $\Pr(X_i=0)=1-p_i$

X	$=\min(X_1, \ldots, X_n)$
T_i	lifetime of a device or equipment
$R_i(t)$	reliability of the device i at time t $(=\Pr\{T_i > t\})$
$r_i(t)$	hazard rate $(=-R_i'(t)/R(t))$
$R(t)$	reliability of a system consisting of n independent components $1, 2, \ldots, n$
$1/a_i$	mean of exponential distribution
$Z(t)$	accumulated damage by time t

APPENDIX 2

TABLE OF LAPLACE TRANSFORMS

Laplace transform of $f(t)$ is defined by

$$L\{f(t)\} = \int_0^\infty e^{-st} f(t)\, dt$$

	$f(t)$	$L\{f(t)\}$
1.	1	$\frac{1}{s}$
2.	t^{n-1}	$\frac{(n-1)!}{s^n}$, $n = 1, 2, 3, \ldots$
3.	t^{a-1}	$\frac{\Gamma(a)}{s^a}$, $a > 0$
4.	$\frac{1}{\sqrt{(\pi t)}}$	$\frac{1}{\sqrt{s}}$
5.	e^{at}	$\frac{1}{s-a}$
6.	$\frac{t^{n-1} e^{at}}{(n-1)!}$	$\frac{1}{(s-a)^n}$, $n = 1, 2, \ldots$
7.	$\sin at$	$\frac{a}{s^2 + a^2}$
8.	$\cos at$	$\frac{s}{s^2 + a^2}$
9.	$f(t) = 0, 0 < t < a$ $= 1, a < t$ (Heaviside unit function)	$\frac{e^{-as}}{s}$
10.	$\delta(t - a)$ (Dirac δ function located at a)	e^{-as}
11.	$\lambda e^{-\lambda t}$ $(\lambda > 0)$	$\frac{\lambda}{s + \lambda}$
12.	$\frac{\lambda^k t^{k-1} e^{-\lambda t}}{\Gamma(k)}$ $(\lambda > 0, k > 0)$	$\left[\frac{\lambda}{s + \lambda}\right]^k$

13.	$\frac{a \exp(-a^2/2t)}{\sqrt{(2\pi t^3)}}$	$\exp(-a\sqrt{(2s)})$
14.	$a^n I_n(at)$	$\frac{[s - (s^2 - a^2)]^n}{\sqrt{(s^2 - a^2)}}, n > -1$
15.	$\frac{n}{t} I_n(at)$	$[s - \sqrt{(s^2 - a^2)}]^n, n = 1, 2, \ldots$
16.	$\{\sqrt{(t^n)}\} I_n(2\sqrt{t})$	$\frac{e^{1/s}}{s^{n+1}}, n = 1, 2, \ldots$

APPENDIX 3

IMPORTANT PROBABILITY DISTRIBUTIONS AND THEIR PARAMETERS, MOMENTS, GENERATING FUNCTIONS

[illegible]	Probability function	Variance	Probability gene-	Laplace transform	Characteristic

Subject Index

Abel's lemma, 66
Addition of random variables, 5, 7, 22, 23
Additive process, 160
Age-dependent branching process, 267–271, 274
  definition of —, 267
  expectation of —, 269
  extinction probability of —, 271
  generating function of —, 268
  multi-type —, 271
Ancestor, 241, 318
Arrival epoch system size, 324
Arrival rate, 278
Autocorrelation function, 231 (see correlation function)
  —generating function, 231–232
Autocovariance function, 106, 231 (see covariance function)
  —generating function, 231–232
Autoregressive integrated moving average process (ARIMA), 236
Autoregressive moving average process (ARMA), 229
Autoregressive process, 226
  —of order one, 233
  —of order two, 228
Availibility, 342, 349

Basic limit theorem of renewal theory, 66
Bellman-Harris process (see age-dependent branching process)
Bernoulli process, 45
Bessel function, 40
  —density, 40
Binary splitting process, 273
Biology, 122, 272
Biorthogonal system, 34
Birth and death process, 120–122, 143, 265, 294–295
Birth and death rates, 122
Bochner-Khinchine theorem, 231, 235
Box-Jenkins method, 236
Branching process, 92, 241–276, 319
  age-dependent—(Bellman-Harris process), 267–271
  continuous time Markov —, 262–267
  continuous time multi-type —, 271
  extension and generalisation of continuous time —, 271–272
  Galton-Watson, —, 241–262
  general —, 271
  Markov renewal —, 330
  multi-type age-dependent —, 271
  multi-type Bellman-Harris —, 271
  Sevast'yanov type —, 271
  —with disasters, 272
Brownian motion, 146
  —process, 146, 148
Busy cycle, 197
Busy period, 197 290, 317
  — and branching process, 318
  number served in a —, 320
  — of *M*/*G*/1 model, 317, 320
  — of *M*/*M*/1 model, 290–293, 320
  — of $M/M^{a,b}/1$ model, 346

Central limit theorem, 146
Chapman-Kolmogorov equations, 60–62, 85, 101, 127–129, 147, 262, 264, 314, 336
  — for Markov chain (with continuous state space), 85
  — for Markov chain (with discrete state space), 60–62
  — for Markov process (with continuous state space), 147
  — for Markov process (with discrete state space), 127–129
Characteristic function, 2
Classical ruin problem, 27
Classification of Markov chains, 63
  aperiodic, 70, 74, 75, 77, 92
  denumerable, 60, 72, 76
  ergodic, 65, 74, 75
  homogeneous, 61, 79, 85, 90
  irreducible, 64, 68, 69, 74–78, 92
  periodic, 72
  reducible, 64
Classification of states of a Markov chain, 63
  absorbing states, 64
  aperiodic states, 65, 67, 68

- ergodic, 65
- non-null persistent state, 65–69, 74–77
- null persistent states, 65–70
- periodic states, 65, 67, 68, 69
- transient states, 65–69, 77

Cobb-Douglas production function, 302
Collective risk theory, 114
Communication system, 62, 90
Conditional expectation, 9
Conditional limit laws of G.W. branching process, 253–258, 272
Conservation method, 280
Conservation principle, 280
Consumer brand switching model, 93
Continuity theorem of Laplace transforms, 21
Continuous time Markov branching process, 262–267, 269
- definition of —, 263
- generating function of —, 263
- probability of extinction of —, 266

Continuous time Markov chains, 126–137
Convolution, 6, 21
- — theorem of Laplace transforms, 21

Correlogram, 224
Correlation function, 106, 224, 227
- properties of functions, 230

Counting process, 97
Covariance function, 48, 50, 224, 230, 234
- properties of functions, 230

Covariance stationary (process), 48, 50, 51, 63, 224, 226, 228, 230
Cyclic component, 223

Demography, 272
Departure epoch system size, 313, 329
Difference equations, 24–30, 41
- characteristic equation of —, 26, 27
- differential —, 30–33, 42
- homogeneous —, 25
- — in probability theory, 27–30
- non-homogeneous —, 29

Diffusion coefficient, 151
Diffusion equations, 147, 149
- forward —, 149
- backward —, 149

Diffusion processes, 146
Dirac function, 17
Direct descendants, 241, 318
Directed graph, 77
- arc-weight of an intree, 137
- intree to a point in a —, 78
- weight of an intree, 78

Distribution,
- Bernoulli —. 39, 331
- binomial —, 4
- chi-square —, 20, 81–83
- compound Poisson —, 9, 39
- decapitated (modified) geometric —, 38
- Erlang —, 20, 22, 138–139
- exponential —, 4, 17, 18, 22, 24, 40
- gamma —, 19, 22
- inverse Gaussian —, 155, 156
- geometric —, 3, 7, 38, 39, 113, 186
- geometric and exponential —, 111
- lattice —, 176
- mixed exponential, 167, 178, 200
- modified exponential —, 41
- multivariate normal —, 49
- negative binomial —, 7, 39, 113, 162, 186
- negative exponential —, see exponential—
- non-lattice —, 176
- normal —, 49
- Pascal —, 7
- Poisson —, 3, 6, 17, 39, 40
- special kind of discrete —, 17
- triangular —, 41
- uniform (rectangular) —, 18, 41
- zero-truncated (decapitated) Poisson, 39, 40, 116

Distribution of total number of progeny, 252–253, 273
Dynamic indeterminism, 45

Ecology, 122
Economics, 122, 146
Eigenvalue, 33
Eigenvector, 33
- left —, 34
- right —, 34

Entrance probabilities, 220
Epidemiology, 272
Erlangs, 278
Erlang's first formula, 299
Erlang's loss formula —, 135, 137, 299
Erlang process, 138
Euler's constant, 335
Extinction of families, 241

Fatou's lemma, 74
Fibonacci numbers, 27, 38
First passage (time) distribution, 65, 152, 209, 293
Flow graph formula, 77
Fokker-Planck equation, 151
Fourier-transform, 2
Frequency domain, 230

Functional equation technique, 272

Galton-Watson branching process, 241–262
  conditioned limit laws of —, 253–258, 272
  critical —, 244, 253
  definition of —, 241
  generalisation of the classical —, 258–260
  generating function of —, 242–244
  — in varying and random environments, 260–261
  moments of —, 243–244
  multitype —, 261–262
  probability of extinction of —, 244–248
  subcritical —, 244, 256
  supercritical —, 244
  time to extinction in a —, 274
  total number of progeny in —, 252–253
  — with immigration, 258–260
Gamma process, 161, 162
Gaussian process, 49
General (random) epoch system state, 326
General branching process, 271
Generating functions, 1–12, 29
  bivariate probability —, 10, 40
  factorial moment —, 2
  moment —, 2
  — of bivariate distribution, 9
  — of branching processes, 242–244, 262
  — of renewal processes, 181–182
  — of sums of random variables, 5–9
  probability —, 1–12, 42
Graph theoretic approach, 77, 136, 144
Graph theoretic formula for finding stationary distribution
  — for Markov chain, 77–79
  — for continuous time Markov chain, 136–137

Hazard rate, 333
  — average, 333
  increasing — (IHR), 333, 341
  increasing — average (IHRA), 333, 341

Heat equation, 150
Heavy traffic, 330, 331

Idle period, 197, 292
Immigration-death process, 126
Immigration-emigration process, 123
Input, 277
Integrated spectrum, 231
Interarrival time, 107, 108, 164, 277
  — in a renewal process, 164
Intermittently available server, 305
Interval transition probability matrix, 211
Inverse Gaussian process, 161
Inventory model, 92

Jirina process, 272

Key-Renewal theorem, 176, 210, 270

Ladder processes, 331
Laplace transform, 12–24, 41, 42
  continuity theorem of transforms, 21
  convolution theorem of transforms, 21
  inverse —, 14
  — of Dirac function, 17
  — of probability distributions, 15–24
  properties of transforms, 13–14
  uniqueness theorem of transforms, 21
Laplace-Stieltje's transform, 25
Lattice random variable, 176
Lifetime distributions, 333, 334
Likelihood ratio criteria, 81
Linear growth process, 123–125, 141, 142
  extinction probability in —, 125
  —with immigration, 126, 142
Little's formula, 280, 284
Loss and delay system, 344
  combination of —, 344
Loss function, 83
Loss system, 134, 277, 299
Lotka's Model, 250

MAICE, 82–83
Machine interference problem, 135, 302, 345

Maitainability, 342
Malthusian parameter, 271
Markov branching process, see continuous time Markov branching process
Markov chain, 48, 54–95
  Chapman-Kolmogorov equation of —, 60–61, 62, 85
  classification of chains, 63–64
  classification of states of chains, 63
  continuous time chains, 126–137
  definition of —, 56
  denumerable (or denumerably infinite) —, 60, 72
  determination of the order of —, (by MAICE), 82–83

general existence theorem of chains, 58
graph theoretic approach for finding stationary distribution of a —, 77–78
grouping of states of a —, 92
higher transition probabilities of a —, 60, 69
homogeneous —, 56, 61
homogenous (stationary) transition probabilities of a —, 56
hypothesis testing in —, 81–82
irreducible —, 64, 68, 69, 74–78, 92
limiting behaviour of —, 70, 74
maximum likelihood estimation of transition probabilities of a—, 79–82
nonstationary (nonhomogenous) chains—, 56, 87–88, 90
— of order one, 58, 82
— of order zero, 8
order of a —, 58, 72, 81
state space of chains, 56
stationary (invariant) distribution of a —, 73, 74, 75
statistical inference for chains, 79–84
three state —. 71, 72, 75
transition probability matrix of a —, 54
two state —, 71, 75
— with continuous state space, 84–87
Markov process, 47, 51, 126, 146, 214, 215
  definition of , 48
  ergodicity of homogeneous —, 133
  three-state —, 144
  two-state —, 131, 134, 349
  — with continuous state space, 146–163
  — with discrete state space, 96, 126–137
Markov renewal equation, 210–211
Markov renewal function, 209–210
Markov renewal process, 195, 203–222, 326, 342
  definition of —, 204, 208
  first passage time in —, 209, 216–218, 220
  interval t.p.m. of —, 211
  recurrent —, 210
  states of —, 210
  theory of birth as a —, 204
  transient —, 210
Markov system, 73
Markov time (stopping time), 170
Matrix, 33
  causative, 87–90
  characteristic equation of a —, 33
  characteristic (latent) roots of —, 33
  constituent matrices, 36
  diagonal —, 36, 70
  doubly stochastic —, 37, 43, 92
  eigenvalues (spectral values) of a —, 33, 34, 38, 43
  eigenvectors of a —, 33, 34, 43, 70
  imprimitive (cyclic) —, 35
  irreducible —, 35, 38, 64
  non-negative matrices, 35
  permutation matrices, 35, 43
  primitive matrices, 35, 38, 43, 70–72
  reducible —, 35
  spectral radius of a —, 33
  spectral representation of a —, 36, 72
  spectral theorem of —, 70, 71, 129
  spectrum of a —, 33
  stochastic (Markov) —, 37, 38, 57
  trace of a —, 34
Meteorology, 223
Models of time series, 224
  autoregressive process, 226, 228
  autoregressive moving average process (ARMA), 229
  continuous parameter process, 234
  first order Markov process, 233
  integrated autoregressive and moving average process (ARIMA), 236
  moving average (MA) process, 225
  non-stationary process, 236
  white noise process, 224

Negative binomial process, 162
New Better than used (NBU), 181, 200
  — in expectation (NBUE), 181
Newton-Raphson method, 325
New worse than used (NWU), 181
  — in expectation, (NWUE), 181
Noise, 223
Non-Markovian chain, 62, 92
Nuclear physics and chemistry, 272
Number of renewals in a random time, 185, 200
  average —, 170

Occupancy problem, 91
Offspring distribution, 241
One-minus-one process, 62
Operational time, 119
  — interval, 119
Operator
  backward shift —, 225
  displacement —, 25

Ornstein-Uhlenbeck process, 156–158, 161
  — as a transformation of Wiener process, 159
Output process, 343

Perron-Frobenius theorem, 35
Poisson cluster process, 113–114, 141
Poisson counting process, 112, 186
Poisson increment process, 52
Poisson process, 49, 96–120, 130, 138, 139
  — and related distributions, 107–110
  autocorrelation function of —, 106
  characterisation of —, 107
  compound —, 106, 113–115, 341, 342
  decomposition of a —, 106
  deterministic selection of —, 139
  doubly stochastic —, 120
  estimation of the parameter of a —, 101
  generalisation of —, 113–120
  — in higher dimensions, 113
  interarrival time of a —, 107, 108
  linear combination of independent processes, 114
  mixed —, 162
  non-homogeneous (time-dependent) —, 101, 118, 119, 342
  postulates for —, 97–98
  properties of processes —, 102–106, 110–112
  random selection of —, 139
  random variation of the parameter of —, 119
  waiting time of —, 108
Pollaczek-Khinchine formula, 313, 314–315
  —mean value formula, 317
  — transform formula, 315, 317
Polya process, 119
Polya's urn model, 54, 55
Power spectrum, 230
Probability of (ultimate) extinction or extinction probability,
  — in a continuous time branching process, 266–267
  —in a G.W. branching process, 244–248
  — in a linear growth process, 125, 142
  — in an age-dependent branching process, 271
Probability vector, 43
Progeny, 252, 318
Programming procedure, 79

Pure birth process 116, 141
Pure death process, 126, 141
Purely random event, 109

Queue discipline, 277
Queueing models,
  $D/M/1$, 326
  $E_k/M/1$, 325
  $GI/M/1$, 323–325
  $M/D/1$, 314
  $M/E_k/1$, 313
  $M/G/1$, 208, 311–313
  $M/G^{a,a}/1$, 347
  $M/G^{a,b}/1$, 327–328
  $M/G/\infty$, 321–322
  $M/G/s/s$, 345
  $M^{(X)}/G^{a,b}/1$, 347
  $M/M/1$, 206, 215, 280–293
  $M/M/s$, 295–298
  $M/M/\infty$, 279, 289–299, 320–322
  $M/M/1/R$, 284–286
  $M/G/s/s$, 300
  $M/M/s/s$, 299–300
  $M/M^{1,b}/1$, 329
  $M/M^{a,b}/1$, 305–310, 329
  $M/M^{a,b}/s$, 347
  $M/M^{a,a}/1$, 347
Queueing processes, 86, 197, 278
  non-birth and death —, 303
Queues,
  average cost in batch service —, 348
  bulk —, 303
  bulk service —, 304–310
  centralisation of counters in —, 343
  combination of loss and delay system in —, 344
  group arrival —, 303
  martingale approach in —, 331
  multichannel —, 294–300
  multistage Poisson queueing system, 346
  optimal design and control in —, 331
  Wiener-Hopf techniques in —, 330
  — with finite waiting space, 284–286
  — with limited input source, 301–302
  — with limited waiting time, 300
  zero-avoiding transitions in —, 343

Random function (or process) see stochastic process
Random geometric sum of exponential variables, 24
Random modification (or residual time), 110
Random selection, 139

Random transmission
  (semi-binary), 52
Random variables, 1, 5, 7, 22, 23, 176
  integer-valued —, 1
  lattice —, 176
  sums of —, 5, 7, 22, 23
Random variation of the parameter of a
  Poisson process, 119
Random walk, 59, 147
  continuous limit of simple —, 147
  — having continuous r.v. as steps, 85
  — with absorbing barriers, 59
  — with reflecting barriers, 76
Recurrence time, 65, 187–189
  backward —, 187–188
  forward —, 187–189
Reflection principle, 152
Regeneration point, 164, 311, 326
  system state at —, 326
Regenerative stochastic process, 195, 196
  existence of limits of —, 195
Reliability, 332–342
  application of $M/M/\infty$ queueing
    model in —, 322
  k-out-of n system —, 332
  Markovian model in —, 336
  — optimisation, 342
  parallel system, 332, 338
  parallel system with maintenance, 338
  series system —, 332
  shock models in —, 339–342
  system —, 334, 335
  system reliability in random environment, 349
  — theory, 122, 203
  wear processes in—, 339–342
Renewal argument, 168
Renewal (or regeneration) epoch, 164
Renewal equation, 167–169
  solution of —, 169
Renewal function, 165
  L.T. of —, 165
  — of age and block replacement
    policies, 174
  — of equilibrium renewal process,
    179–180
Renewal process, 120, 164–202
  alternating (or two-stage) —, 192–195
  definition of —, 164
  delayed (modified or general) —, 177–
    179, 189, 193
  equilibrium (stationary) —, 179
  interarrival time in a —, 164, 166
  ordinary —, 177, 178, 216
  p g.f. of processes —, 181
  renewal rate, 179
  uniform —, 189
Renewal reward process, 190–192, 348
  age-based —, 191, 201
  definition of, 190
Renewal theorems, 169–177, 216
  Blackwell's —, 176, 210
  elementary —, 170, 172
  key (Smith's) —, 176, 210
Renewal theory, 164
  application to a management problem,
    183
  basic limit theorem of —, 66
  generalisation of the classical —,
    198–199
  integral equation of —, 168
  — in queueing process, 197
Renewal type equation 169, 188
Replacement interval, 174
Replacement policies, 173–175
  age —, 173
  age-based renewal processes under—,
    191
  block —, 174, 202
  number of removals under —, 200
  renewal functions under —, 174
  renewal processes under —, 174, 175
Residual and excess (or spent) life times,
    186–187, 190, 216
Rouchè's theorem, 290, 306

Sample paths (or trajectories),
  — of a stochastic process, 152
  — of a Wiener process, 152
Schroder functional equation, 258
Seasonal components, 224
Semi-Markov process, 195, 203–221
  definition of —, 204
  embedded Markov chain of —, 204
  — in queues, 206–208, 215, 221, 313
  limiting distribution of —, 212,
    215–216
  recurrence times of —, 215–216, 220
  waiting time in —, 205
Service time, 278
Shock models, 339–342
Signal, 223
Spares inventory model, 322
Spectral density function, 231
  estimation of —, 237
  normalised —, 231
Stability of a Markov system, 73
State space, 46
  discrete —, 46, 47, 56, 84, 85
  continuous —, 46, 47, 84-85

Stationary process, 48
  covariance, —, 48
Steady state distribution, 279
Stochastic difference equation, 225
Stochastic processes, 45
  covariance function of —, 48–51
  covariance stationary (weakly or wide sense stationary), 48–53
  continuous parameter, 47
  discrete parameter, 47
  evolutionary, 49, 50
  Gaussian, 49
  multi-dimensional, 46, 47
  one-dimensional, 46, 47
  one-minus-one, 52
  specification of, 46
  state space of, 46
  strictly stationary, 48, 49, 51
  — with independent increments, 47
Stochastic sequence, 47
Stopping time, 170, 171, 172
Supplementary variable technique, 330
System state
  arrival point —, 324
  departure point —, 311, 326
  general time —, 313, 326, 329

Tackacs integral equation, 317–318
Tauberian theorems, 14
Telegraph signals
  random —, 52
  semi-random —, 52
Telephone traffic, 277
Threshold, 341, 342
Time domain, 230
Time series, 223–240
  continuous parameter —, 234
  models of —, 224-230
  statistical analysis of —, 235
  statistical inference in —, 235
Traffic intensity, 278
Transient behaviour,
  — of $M/M/1$ queue, 286–293
  — of $M/M/1/1$ queue, 302-303
  — of $M/G/\infty$ queue, 320–322
Transition density, 127, 128, 137
  — of continuous time branching process, 263-264
  — of continuous time Markov chain, 127, 128
  — (rate) matrix, 128, 135, 144
Transition probability, 56
  higher—, 60
  — matrix, 57
Trend, 223
Tshebyshev's lemma, 102

Uniqueness theorem of Laplace transforms, 21

Virtual waiting time, 293

Waiting time, 108, 205, 282
  — for a change of state, 127
  — for the first success, 4
  — process, 293
  virtual —, 293
Waiting time distributions, 282
  — in $GI/M/1$ queue, 325
  — in $M/M/1$ queue, 282--284
  – in $M/M^{a,b}/1$ queue, 307–310
  — in $M/M/s$ queue, 296-297
Wald's equation, 170–171
Weakly (wide-sense) stationary, 48
Wear proeess, 339–342
White noise process, 224
Wiener-Hopf technique, 330
Wiener-Einstein procees, 147 (see Wiener process), 147–156, 159
  backward diffusion equation of —, 149
  definition of, 148
  differential equations for a —, 149
  distribution of maximum of —, 142
  drift of —, 148
  first passage time distribution of —, 152–156
  forward diffusion equation of —, 149
  Kolmogorov equation of —, 150
  sample path of —, 152

Yaglom's theorem, 256, 259, 260
Yule process, 228
Yule-Furry process, 116-117, 142, 143
Yule-Walker equations , 228

Zero-avoiding state probability, 290
Zero-truncated Poisson distribution, 39,. 40, 116
Zipf's rank size law, 141

Author Index

Anderson, T.W., 79, 81, 94, 224, 236, 239
Arora, K.L., 310, 351
Aspden, P., 353
Athreya, K.B., 93, 94, 199, 261, 262, 269, 271, 275

Bailey, N.T.J., 42, 43, 44, 145, 286, 305, 351
Balkema, A.A., 189, 202, 351
Barlow, R., 173, 175, 200, 202, 322, 342, 351
Barndorff-Nielson, 162
Barrer, D.Y., 351
Bartholomew, D.J., 203, 221
Bartlett, M.S., 79, 120, 145, 237, 239
Bellman, R.E., 30, 44, 267, 275
Belyayev, B.V., 351
Bharucha-Reid, A.T., 94, 145
Bhat, U.N., 203, 221, 305, 351, 355
Billingsley, P., 79, 94, 331
Birnbaun, Z.W., 333, 351
Blackman, R.B., 237, 239
Blackwell, D., 176, 198
Blanc-Lapierre, A., 145, 151, 162
Bloemena, A.R., 305, 351
Bloomfield, P., 237, 239
Borel, E., 293, 351
Borthakur, A., 310, 330, 347, 351
Box, G.E.P., 236, 239
Bretherton, M.H., 142
Brockmeyer, E., 277, 352
Brumelle, S.L., 280, 300, 352
Burke, P.J., 343, 352

Champernowne, D.C., 286, 352
Chan, J., 302, 352
Chatfield, C., 236, 239
Chaudhry, M.L., 310, 330, 352
Cheong, C.K., 219, 222
Chhikara, R.S., 155, 163
Chow, Y.S., 198, 199
Christer, A.H., 191, 192, 202
Chung, K.L., 58, 94, 127, 198
Cinlar, E., 203, 216, 219, 222, 342
Clarke, A.B., 331, 352
Cohen, J.W., 313, 352
Cohn, H., 90, 95
Collin, L., 94
Conolly, B.W., 33, 44, 286, 302, 352
Cooke, K.L., 30, 44
Cox, D.R., 94, 145, 155, 162, 202, 302, 331, 339, 352
Crabill, T.B., 331, 352
Cramer, H., 119, 145
Cromie, M.V., 310

Daley, D.J., 352
Davies, N., 226, 239
De Haan, L., 202, 351
De Smit, J.H.A., 222
Doob, J.L., 53, 162
Downton, F., 305, 352

Einstein, A., 146, 163
Enns, E.G., 320, 352
Erlang, A.K., 277, 278
Esary, J.D., 333, 339, 353

Fabens, A.J., 222, 305, 330, 347, 353
Fearn, D.H., 260, 275
Feller, W., 44, 62, 66, 94, 145, 169, 170, 177, 198, 202, 302, 319
Folks, J.L., 155, 163
Fortet, R., 145, 151, 162, 300, 353
Foster, J.A., 260
Frost, M.G., 239
Fuller, W.A., 224, 236, 238, 239

Gabriel, K.R., 83
Galton, F., 241, 244, 272, 275
Gaver, D.P., 40
Ghare, P.M., 310, 353
Girault, M., 145
Gnedenko, B.V., 338, 351, 353
Good, I.J., 319
Goldberg, R.R., 225, 239
Goodman, L.A., 79, 81, 94
Grandell, J., 120, 145
Granger, C.W.J., 236, 239
Grenandar, U., 237, 239
Gross, D., 287, 296, 313, 352, 353
Gumbel, E.J., 298

25(45-95/1982)

Hamdan, M.A., 40
Hammed, A., 342, 351
Hannan, E.J., 237, 239
Harary, F., 87, 94
Harris, C.M., 287, 296, 313, 353
Harris, T.E., 249, 267, 269, 275
Hatanka, M., 239
Heathcote, C.R., 259, 260, 275
Heyde, C.C., 199
Howard, R.A., 222
Hwang, C.L., 354

Iosifescu, M., 58, 163
Isaacson, D.L., 90, 95

Jackson, R.R.P., 346, 353
Jagers, P., 260, 271, 272, 275
Jaiswal, N.K., 305, 353
Jenkins, G.M., 236, 237, 239, 240
Joffe, A., 258, 275
Johnson, N.L., 155, 163
Jordan, C., 44, 358
Judge, G.G., 95

Kao, E.P.C., 203, 222
Kaplan, N., 261, 272, 275, 276
Karlin, S., 63, 95, 110, 145, 163, 261, 275, 295, 353
Karmarkar, U.S., 178, 183, 202
Kaufman, J.S., 345, 353
Keilson, J., 311, 353
Kemeny, J.G., 95
Kendall, D.G., 279, 311, 353
Kendall, M.G., 224, 239
Kesten, H., 199
Keyfit, N., 204, 222, 272, 275
Kingman, J.F.C., 90, 95, 112, 120, 145, 325, 330, 353
Kleinrock, L., 316, 353
Kooharian, A., 311, 353
Koopmans, L.H., 236, 240
Kosten, L., 300, 353
Kotz, S., 155, 162
Kovalenko, I.N., 353
Krakowski, M., 280, 353
Kuo, W., 355

Lambotte, J.P., 355
Lancaster, P., 44
Lee, T.C., 79, 95, 353
Leiter, R,E., 40
Lemoine, A.J., 316, 353
Levhari, D., 302, 353
Lévy, P., 163, 203, 222
Lie, C.H., 342, 354
Lilliefors, H.W., 331
Lipstein, B., 94
Little, J.D.C., 280, 354
Loris-Teghem, J., 355

Madsen, R.W., 90, 95
Magazine, M.G., 352
Malinvaud, B., 224, 240
Marshall, A.W., 333, 339, 351, 353, 354
Mcgregor, J., 295, 353
Medhi, J., 39, 83, 95, 237, 239, 240, 307, 310, 330, 351, 354
Mehata, K.M., 276
Menken, J.A., 222
Miller, H.D., 94, 145, 155, 162
Mode, C.J., 261, 262, 272, 276
Molina, E.C., 354
Morris, M.J., 239
Morse, P.M., 280, 354

Nakagawa, T., 350, 354
Naor, P., 302, 354
Natvig, B., 354
Neumann, J., 83
Neuts, M., 222, 290, 305, 310, 330, 347, 354
Ney, P., 93, 94, 262, 269, 271, 275
Neyman, J., 45, 63

Ord, K., 141, 145
Ornstein, L.S., 157

Pakes, A.G., 272, 276
Parzen, E., 44, 62, 95, 138, 145, 237, 240
Pata, M.B., 239
Paz, A., 90, 95
Pollaczek, F., 300
Potter, R.G., 204, 222
Prabhu, N.U., 53, 63, 66, 95, 145, 161, 163, 177, 198, 202, 286, 293, 296, 320, 330, 354, 355
Priestley, M.B., 236, 237, 240
Proctor, C.L., 144, 145
Prohorov, Yu.V., 154, 163
Proschan, F., 173, 175, 200, 202, 322, 339, 342, 351, 353
Pyke, R., 203, 216, 222

Rade, L., 350, 355
Rao, T. Subba, 237, 240
Reich, E., 355
Robbins, H., 198
Rosenblatt, M., 239
Ross, S.M., 202, 222
Rozanov, Yu.A., 154, 163

Rubinovitch, M., 355
Runnenberg, J.T., 315, 355
Ryan, T.A., Jr., 199

Saaty, T.L., 286
Sankaranarayanan, G., 199
Scott, M., 320, 355
Seneta, E., 90, 95, 260, 272, 276
Serfozo, R.F., 120
Sevast'yanov, B., 271, 276, 300
Shaked, M., 339, 354
Sheps, M., 204, 222
Shesinski, E., 302, 353
Shimshak, D.G., 355
Shonick, W., 344, 355
Sim, S.H., 310
Sing, B., 144, 145
Sirjaev, A.N., 170, 202
Smith, M.G., 29, 44
Smith, W.L., 170, 176, 199, 202, 203, 222, 260, 276
Snell, J.L., 95
Sobel, M.J., 331, 355
Solberg, J.J., 77, 95, 137
Solovyev, A.D., 351
Sphicas, G.P., 355
Srinivasan, S.K., 272, 276
Stidham, S., 280, 331, 355
Stuart, A., 224, 239
Styan, G.P.H., 94
Subelman, E.J., 95

Takacs, L., 300, 303, 305, 317, 355
Tautu, P., 58, 163
Teghem, J., 310, 330, 355
Templeton, J.G.C., 310, 330, 352
Teugels, J.L., 222
Tillman, F.A., 342, 354, 355
Tong, H., 82, 95, 240
Torrez, W.C., 276
Tukey, J.W., 237, 239

Uhlenbeck, G.E., 157
Ulmer, M.B., Jr., 355

Wang, C.L., 355
Wasan, M.T., 161, 163
Watson, H.W., 241, 272, 275
Watts, D.G., 237, 240
Weiss, H.J., 310, 348, 355
Whitaker, D., 93, 95
Whitt, W., 330, 355
Whittle, P., 79, 95
Widder, D.V., 44
Wilkinson, W.E., 260, 261, 276
Williamson, E., 142
Wold, H., 226, 237, 240
Wolff, R.W., 331, 355

Yang, Y.S., 260, 276
Yeo, G.F., 162

Zellner, A., 95

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

STOCHASTIC PROCESSES

519.2
MED

Aims at the level between elementary texts and advanced stochastic processes. The pre-requisites are a course of ele probability theory and statistics, and a course of advanced The theoretical results developed have been followed by a larg number of illustrative examples, both theoretical and applied. These have been supplemented by a large number of exercises, answers to most of which are given. The book will be suitable as a text for advanced undergraduate, postgraduate and research level courses in Statistics, Applied Mathematics, Operations Research, Computer Science, Industrial Engineering etc. It will also be useful to such other readers who want to apply the theory of stochastic processes to their professional or research work.

21C

Jyotiprasad Medhi

Professor J Medhi is currently Senior Professor & Head Department of Statistics, Gauhati University, Gauhati, India. He earned his Master's degree in Pure Mathematics from the University of Calcutta, India, Master's degree in Statistics (by thesis) from the University of Manchester, U.K. and his doctorate in Statistics (Stochastic Processes) from the University of Paris, France in 1956. Since 1949 he has been a Faculty member of the Post-Graduate Department of Mathematics & Statistics, Gauhati University. He was a Visiting Professor at the University of Montreal, Canada in 1967-68, and was also on a visiting assignment at the University of Toronto, Canada during Fall 1982. In 1979 he visited France and U.K. on a study-cum-lecture tour and for exchange of ideas with academics in those countries.

His first research paper appeared in 1955. He has since published numerous papers in such International Journals as *Biometrika*, *Skand. Akt.*, *Metrika*, *Management Science*, *Cahiers du Centre d'etude de Recherche Operationnelle*, *Sankhya*, *J. Ind. Stat. Ass.*, *Ind. J. of Met. & Geophysics* etc. His current research interests are Applied Probability and Stochastic Processes (with special reference to Queueing Processes) and Operations Research.

Professor Medhi is also Director of Population Research Centre set up by the Ministry of Health & Family Welfare, Government of India in his Department. He has been a member of the Institute of Mathematical Statistics, U.S.A. since 1963 and was, President, Section of Statistics, Indian Science Congress Association 1978-79.

WILEY EASTERN LIMITED
New Delhi Bangalore Bombay Calcutta Madras Hyderabad